(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 March 2001 (01.03.2001)

PCT

(10) International Publication Number
WO 01/14550 A1

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, C12Q 1/68, C07K 16/18, A61K 67/027

(21) International Application Number: PCT/IB00/01098

(22) International Filing Date: 28 July 2000 (28.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,941 19 August 1999 (19.08.1999) US

(71) Applicant *(for all designated States except US)*: **GENSET** [FR/FR]; Intellectual Property Department, 24, rue Royale, F-75008 Paris (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BARRY, Caroline** [FR/FR]; Residence le Jardin des Lys, 2, avenue des Cevennes, F-91940 Les Ulis (FR). **CHUMAKOV, Ilya** [FR/FR]; 196, rue des Chevrefeuilles, F-77000 Vaux-le-Penil (FR). **BLUMENFELD, Marta** [FR/FR]; 5, rue Tagore, F-75013 Paris (FR).

(74) Common Representative: **GENSET**; Intellectual Property Department, 24, rue Royale, F-75008 Paris (FR).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*
— *With amended claims.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/14550 A1

(54) Title: PROSTATE CANCER-RELASED GENE 3 (PG-3) AND BIALLELIC MARKERS THEREOF

(57) Abstract: The invention concerns the genomic sequence and cDNA sequences of the PG-3 gene. The invention also concerns biallelic markers of the PG-3 gene. The invention also concerns polypeptides encoded by the PG-3 gene. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents.

PROSTATE CANCER-RELATED GENE 3 (PG3) AND BIALLELIC MARKERS THEREOF

FIELD OF THE INVENTION

The present invention is directed to polynucleotides encoding a PG-3 polypeptide as well as the regulatory regions located at the 5'- and 3'-ends of said coding region. The invention also relates to polypeptides encoded by the PG-3 gene. The invention also relates to antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the PG-3 gene useful in genetic analysis.

BACKGROUND OF THE INVENTION

Cancer is one of the leading causes of death in industrialized countries. This makes cancer a serious burden in terms of public health, especially in view of the aging of the population. Indeed, over the next 25 years there will be a dramatic increase in the number of people developing cancer. Globally, 10 million new cancer patients are diagnosed each year and there will be 20 million new cancer diagnoses by the year 2020.

In spite of a large number of available therapeutic techniques including but not limited to surgery, chemotherapy, radiotherapy, bone marow transplantation, and in spite of encouraging results obtained with experimental protocols in immunotherapy or gene therapy, the overall survival rate of cancer patients does not reach 50% after 5 years . Therefore, there is a strong need for both a reliable diagnostic procedure which would enable early-stage cancer prognosis, and for preventive and curative treatments of the disease.

A cancer is a clonal proliferation of cells produced as a consequence of cumulative genetic damage that finally results in unrestrained cell growth, tissue invasion and metastasis (cell transformation). Regardless of the type of cancer, transformed cells carry damaged DNA as gross chromosomal translocations or, more subtly, as DNA amplification, rearrangement or even point mutations.

Cancer is caused by the dysregulation of the expression of certain genes. The development of a tumor requires an important succession of steps. Each of these comprises the dysregulation of a gene either involved in cell cycle activity or in genomic stability and the emergence of an abnormal mutated clone which overwhelms the other normal cell types because of a proliferative advantage. Cancer indeed happens because of a combination of two mechanisms. Some mutations enhance cell proliferation, increasing the target population of cells for the next mutation. Other mutations affect the stability of the entire genome, increasing the overall mutation rate, as in the case of mismatch repair proteins (reviewed in Arnheim N & Shibata D, 1997).

Recent studies have identified three groups of genes which are frequently mutated in cancer. The first two groups are involved in cell cycle activity , which is a mechanism that drives normal cell proliferation and ensures the normal development and homeostasis of the organism. Conversely, many of the properties of cancer cells - uncontrolled proliferation, increased mutation

rate, abnormal translocations and gen amplifications - can be attributed directly to perturbations of the normal regulation or progression of the cycle.

The first group of genes, called oncogenes, are genes whose products activate cell proliferation. The normal non-mutant versions are called protooncogenes. The mutated forms are excessively or inappropriately active in promoting cell proliferation and act in the cell in a dominant way such that a single mutant allele is enough to affect the cell phenotype. Activated oncogenes are rarely transmitted as germline mutations since they are probably be lethal when expressed in all the cells in the organism. Therefore oncogenes can only be investigated in tumor tissues. Oncogenes and protooncogenes can be classified into several different categories according to their function. This classification includes genes that code for proteins involved in signal transduction such as: growth factors (*i.e., sis, int-2*); receptor and non-receptor protein-tyrosine kinases (*i.e., erbB, src, bcr-abl, met, trk*); membrane-associated G proteins (*i.e., ras*); cytoplasmic protein kinases (*i.e.*, mitogen-activated protein kinase –*MAPK*- family, *raf, mos, pak*), or nuclear transcription factors (*i.e., myc, myb, fos, jun, rel*) (for review see Hunter T, 1991 ; Fanger GR *et al.*, 1997 ; Weiss FU *et al.*, 1997).

The second group of genes which are frequently mutated in cancer, called tumor suppressor genes, are genes whose products inhibit cell growth. Mutant versions in cancer cells have lost their normal function, and act in the cell in a recessive way such that both copies of the gene must be inactivated in order to change the cell phenotype. Most importantly, the tumor phenotype can be rescued by the wild type allele, as shown by cell fusion experiments first described by Harris and colleagues (Harris H *et al.*,1969). Germline mutations of tumor suppressor genes are transmitted and thus studied in both constitutional and tumor DNA from familial or sporadic cases. The current family of tumor suppressors includes DNA-binding transcription factors (*i.e., p53, WT1*), transcription regulators (*i.e., RB, APC,* and *BRCA1*), and protein kinase inhibitors (*i.e., p16*), among others (for review, see Haber D & Harlow E, 1997).

The third group of genes which are frequently mutated in cancer, called mutator genes, are responsible for maintaining genome integrity and/or low mutation rates. Loss of function of both alleles increases cell mutation rates, and as a consequence, proto-oncogenes and tumor suppressor genes are mutated. Mutator genes can also be classified as tumor suppressor genes, except for the fact that tumorigenesis caused by this class of genes cannot be suppressed simply by restoration of a wild-type allele, as described above. Genes whose inactivation may lead to a mutator phenotype include mismatch repair genes (*i.e., MLH1, MSH2*), DNA helicases (*i.e., BLM, WRN*) or other genes involved in DNA repair and genomic stability (*i.e., p53*, possibly *BRCA1* and *BRCA2*) (For review see Haber D & Harlow E, 1997; Fishel & Wilson. 1997 ; Ellis,1997).

The recent development of sophisticated techniques for genetic mapping has resulted in an ever expanding list of genes associated with particular types of human cancers. The human haploid genome contains an estimated 80,000 to 100,000 genes scattered on a 3×10^9 base-long double-

stranded DNA. Each human being is diploid, *i.e.*, possesses two haploid genomes, one from paternal origin, the other from maternal origin. The sequence of a given genetic locus may vary between individuals in a population or between the two copies of the locus on the chromosomes of a single individual. Genetic mapping techniques often exploit these differences, which are called polymorphisms, to map the location of genes associated with human phenotypes.

One mapping technique, called the loss of heterozygosity (LOH) technique, is often employed to detect genes in which a loss of function results in a cancer, such as the tumor suppressor genes described above. Tumor suppressor genes often produce cancer via a two hit mechanism in which a first mutation, such as a point mutation (or a small deletion or insertion) inactivates one allele of the tumor suppressor gene. Often, this first mutation is inherited from generation to generation. A second mutation, often a spontaneous somatic mutation such as a deletion which deletes all or part of the chromosome carrying the other copy of the tumor suppressor gene, results in a cell in which both copies of the tumor suppressor gene are inactive. As a consequence of the deletion in the tumor suppressor gene, one allele is lost for any genetic marker located close to the tumor suppressor gene. Thus, if the patient is heterozygous for a marker, the tumor tissue loses heterozygosity, becoming homozygous or hemizygous. This loss of heterozygosity generally provides strong evidence for the existence of a tumor suppressor gene in the lost region.

LOH has allowed the identification of several chromosomic regions associated with cancer. Indeed, substantial amounts of LOH data support the hypothesis that genes associated with distinct cancer types are located within 8p23 region of the human genome. Several regions of chromosome arm 8p were found to be frequently deleted in a variety of human malignacies including those of the prostate, head and neck, lung and colon. Emi *et al.* demonstrated the involvement of the 8p23.1-8p21.3 region in cases of hepatocellular carcinoma, colorectal cancer, and non-small cell lung cancer (Emi *et al.*, 1992). Yaremko, *et al.*, (1994) showed the existence of two major regions of LOH for chromosome 8 markers in a sample of 87 colorectal carcinomas. The most prominent loss was found for 8p23.1-pter, where 45% of informative cases demonstrated loss of alleles. Scholnick *et al.* (Scholnick *et al*, 1996 and Sunwoo *et al.*, 1996) demonstrated the existence of three distinct regions of LOH for the markers of chromosome 8 in cases of squamous cell carcinoma of the supraglottic larynx. They showed that the allelic loss of 8p23 marker D8S264 serves as a statistically significant, independent predictor of poor prognosis for patients with supraglottic squamous cell carcinoma. The study of 51 squamous cell carcinomas of the head and neck and 29 oral squamous cell carcinoma cell lines showed a frequent allelic loss and homozygous deletion at 1 or more loci located in the 8p23 region (Ishwad CS *et al.*, 1999). In addition, a high resolution deletion map of 150 squamous cell carninomas of the larynx and oral cavity showed two distinct classes of deletion for the 8p23 region within the D8S264 to D8S1788 interval (Sunwoo *et al.*, 1999).

In other studies, Nagai *et al.* (1997) demonstrated the highest loss of heterozygosity in the specific region of 8p23 by genome wide scanning of LOH in 120 cases of hepatocellular carcinoma (HCC). Further studies using high-density polymorphic marker analysis identified three minimal deleted areas on chromosome 8p, one of them being a 5 cM area in 8p23, probably indicative of the presence of a tumor suppressor loci for HCC (Pineau P, *et al.*, 1999). Gronwald *et al.* (1997) also demonstrated 8p23-pter loss in renal clear cell carcinomas.

The same region is involved in specific cases of prostate cancer. Matsuyama *et al.* (1994) showed the specific deletion of the 8p23 band in prostate cancer cases, as monitored by FISH with D8S7 probe. They were able to document a substantial number of cases with deletions of 8p23 but retention of the 8p22 marker LPL. Moreover, Ichikawa *et al.* (1996) deduced the existence of a prostate cancer metastasis suppressor gene and localized it to 8p23-q12 by studies of metastasis suppression in highly metastatic rat prostate cells after transfer of human chromosomes. Recently Washbum *et al.* (1997) were able to find substantial numbers of tumors with the allelic loss specific to 8p23 by LOH studies of 31 cases of human prostate cancer. In these samples they were able to define the minimal overlapping region with deletions covering genetic interval D8S262-D8S277. In addition, using PCR analysis of polymorphic microsatellite repeat markers, 29% of 60 prostate tumors showed LOH, at the locus D8S262 of the 8p23 region (Perinchery *et al.*, 1999).

Recent studies have also implicated the 8p23 region in other types of cancers such as fibrous histiocytomas, ovarian adenocarcinomas and gastric cancers. Indeed, comparative genomic hybridization data showed the involvment of the 8p23.1 region in fibrous histiocytomas and detected a minimal amplified region between D8S1819 and D8S550 containing a gene MASL1, the overexpression of which might be oncogenic (Sakabe *et al.*, 1999). LOH was also observed for 27 ovarian adenocarcinomas on 8p. Detailed examination of nine tumours with partial deletions defined three regions of overlap including two in 8p23 (Wright *et al.*, 1998). Comparative genomic hybridization of 58 primary gastric cancers detected gain of the 8p22-23 region in 24% of the tumors and even high-level amplification of the same region in 5% of the tumors . This amplified region was narrowed down to 8p23.1 by reverse-painting FISH to prophase chromosomes (Sakakura *et al.*, 1999).

The present invention relates to the Prostate Cancer Related Gene 3 or PG-3 gene, a gene present in the 8p23 cancer candidate region, as well as diagnostic methods and reagents for detecting alleles of the PG-3 gene which may cause cancer, and therapies for treating cancer.

SUMMARY OF THE INVENTION

The present invention pertains to nucleic acid molecules comprising the genomic sequence and the cDNA sequence of a novel human gene which encodes a PG-3 protein. The PG-3 gene is localized in the 8p23 candidate region shown to be involved in several types of cancer by LOH studies and presents homology with the *BRCA1* gene involved in transcriptional control through modulation of chromatin structure (Bochar et al, 2000), and in which mutations are thougth to be

responsible for 45% of inherited breast cancer and more than 80% of inherited breast and ovarian cancer. In addition, *BRCA1* carriers have a 4-fold increased risk of colon cancer, whereas male carriers face a 3-fold increased risk of prostate cancer.

The PG-3 genomic sequence comprises regulatory sequences located upstream (5'-end) and downstream (3'-end) of the transcribed portion of said gene, these regulatory sequences being also part of the invention.

The invention also relates to the cDNA sequence encoding the PG-3 protein, as well as to the corresponding translation product.

Oligonucleotide probes or primers hybridizing specifically with a PG-3 genomic or cDNA sequence are also part of the present invention, as well as DNA amplification and detection methods using said primers and probes.

A further object of the invention relates to recombinant vectors comprising any of the nucleic acid sequences described herein, and in particular to recombinant vectors comprising a PG-3 regulatory sequence or a sequence encoding a PG-3 protein. The present invention also relates to host cells and transgenic non-human animals comprising said nucleic acid sequences or recombinant vectors.

The invention further encompasses biallelic markers of the PG-3 gene useful in genetic analysis.

Finally, the invention is directed to methods for the screening of substances or molecules that inhibit the expression of PG-3, as well as to methods for the screening of substances or molecules that interact with a PG-3 polypeptide or that modulate the activity of a PG-3 polypeptide.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of an exemplary computer system.

Figure 2 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database.

Figure 3 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous.

Figure 4 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence.

BRIEF DESCRIPTION OF THE SEQUENCES PROVIDED IN THE SEQUENCE LISTING

SEQ ID No 1 is a genomic sequence of PG-3 comprising the 5' regulatory region (upstream untranscribed region), the exons and introns, and the 3' regulatory region (downstream untranscribed region).

SEQ ID No 2 is a cDNA sequence of PG-3.

SEQ ID No 3 is the amino acid sequence encoded by the cDNA of SEQ ID No 2.

SEQ ID No 4 is a primer containing the additional PU 5' sequence further described in Example 2.

SEQ ID No 5 is a primer containing the additional RP 5' sequence further described in Example 2.

In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to indicate the locations of biallelic markers within the sequences and to identify each of the alleles present at the polymorphic base. The code "r" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is an adenine. The code "y" in the sequences indicates that one allele of the polymorphic base is a thymine, while the other allele is a cytosine. The code "m" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is a cytosine. The code "k" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a thymine. The code "s" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a cytosine. The code "w" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is a thymine. The nucleotide code of the original allele for each biallelic marker is the following:

Biallelic marker	Original allele
5-390-177	C
5-391-43	G
5-392-222	T
5-392-280	T
4-59-27	G
4-58-289	C
4-54-199	A
4-54-180	C
4-51-312	G
99-86-266	A
4-88-107	G
5-397-141	G
5-398-203	C
99-12738-248	A
99-109-358	C
99-12749-175	T
4-21-154	C
4-21-317	G
4-23-326	G
99-12753-34	A
5-364-252	G
99-12755-280	G
99-12755-329	C

4-87-212	A
99-12757-318	C
99-12758-102	G
99-12758-136	C
4-105-98	A
4-105-86	G
4-45-49	T
4-44-277	T
4-86-60	C
4-84-334	G
99-78-321	T
99-12767-36	G
99-12767-143	T
99-12767-189	T
99-12767-380	G
4-80-328	C
4-36-384	C
4-36-264	G
4-36-261	C
4-35-333	A
4-35-240	G
4-35-173	T
4-35-133	C
99-12771-59	T
99-12774-334	A
99-12776-358	G
99-12781-113	A
4-104-298	C
4-104-254	G
4-104-250	C
4-104-214	A
99-12818-289	T
99-24807-271	C
99-24807-84	G
99-12831-157	G
99-12831-241	C
99-12832-387	T
99-12836-30	G
99-12844-262	C
4-24-74	C
4-24-246	C
4-24-314	G

4-27-190	A
5-400-145	G
5-400-149	G
5-400-175	T
5-400-231	T
5-400-367	A
99-12852-110	T
99-12852-325	A
4-37-326	A
4-37-107	G
5-270-92	G
99-12860-47	G
99-12860-57	T
5-402-144	C

In some instances, the polymorphic bases of the biallelic markers alter the identity of an amino acid in the encoded polypeptide. This is indicated in the accompanying Sequence Listing by use of the feature VARIANT, placement of an Xaa at the position of the polymorphic amino acid, and definition of Xaa as the two alternative amino acids. For example if one allele of a biallelic marker is the codon CAC, which encodes histidine, while the other allele of the biallelic marker is CAA, which encodes glutamine, the Sequence Listing for the encoded polypeptide will contain an Xaa at the location of the polymorphic amino acid. In this instance, Xaa would be defined as being histidine or glutamine.

DETAILED DESCRIPTION

The present invention concerns polynucleotides and polypeptides related to the PG-3 gene. Oligonucleotide probes and primers hybridizing specifically with a genomic or a cDNA sequence of PG-3 are also part of the invention. A further object of the invention relates to recombinant vectors comprising any of the nucleic acid sequences described in the present invention, and in particular recombinant vectors comprising a regulatory region of PG-3 or a sequence encoding the PG-3 protein, as well as host cells comprising said nucleic acid sequences or recombinant vectors. The invention also encompasses methods of screening for molecules which inhibit the expression of the PG-3 gene or which modulate the activity of the PG-3 protein. The invention also relates to antibodies directed specifically against such polypeptides that are useful as diagnostic reagents.

The invention also concerns PG-3-related biallelic markers which can be used in any method of genetic analysis including linkage studies in families, linkage disequilibrium studies in populations and association studies of case-control populations. An important aspect of the present invention is that biallelic markers allow association studies to be performed to identify genes involved in complex traits. These biallelic markers may lead to allelic variants of the PG-3 protein.

Definitions

typically comprises about 50%, preferably 60 to 90% weight/weight of a polypeptide or polynucleotide sample, respectively, more usually about 95%, and preferably is over about 99% pure. Polypeptide and polynucleotide purity, or homogeneity, is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art. As an alternative embodiment, purification of the polypeptides and polynucleotides of the present invention may be expressed as "at least" a percent purity relative to heterologous polypeptides and polynucleotides (DNA, RNA or both). As a preferred embodiment, the polypeptides and polynucleotides of the present invention are at least; 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 96%, 98%, 99%, or 100% pure relative to heterologous polypeptides and polynucleotides, respectively. As a further preferred embodiment the polypeptides and polynucleotides have a purity ranging from any number, to the thousandth position, between 90% and 100% (e.g., a polypeptide or polynucleotide at least 99.995% pure) relative to either heterologous polypeptides or polynucleotides, respectively, or as a weight/weight ratio relative to all compounds and molecules other than those existing in the carrier. Each number representing a percent purity, to the thousandth position, may be claimed as individual species of purity.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

As used herein, the term "non-human animal" refers to any non-human vertebrate, birds and more usually mammals, preferably primates, farm animals such as swine, goats, sheep, donkeys, and horses, rabbits or rodents, more preferably rats or mice. As used herein, the term "animal" is used to refer to any vertebrate, preferable a mammal. Both the terms "animal" and "mammal" expressly embrace human subj cts unless preceded with the term "non-human".

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. Antibodies include recombinant proteins comprising the binding domains, as wells as fragments, including Fab, Fab', $F(ab)_2$, and $F(ab')_2$ fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case a PG-3 polypeptide, that determines the specificity of the antigen-antibody reaction. An "epitope" refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by Geysen *et al.* 1984; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (*i.e.* the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule.

As used interchangeably herein, the terms "nucleic acids", "oligonucleotides", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. The term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one of the following modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell required to initiate the specific transcription of a gene.

A sequence which is "operably linked" to a regulatory sequence such as a promoter means that said regulatory element is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the nucleic acid of interest. As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. More precisely, two DNA molecules (such as a polynucleotide containing a promoter region and a polynucleotide encoding a desired polypeptide or polynucleotide) are said to be "operably linked" if the nature of the linkage between the two polynucleotides does not (1) result in the introduction of a frame-shift mutation or (2) interfere with the ability of the polynucleotide containing the promoter to direct the transcription of the coding polynucleotide.

The term "primer" denotes a specific oligonucleotide sequence which is complementary to a target nucleotide sequence and used to hybridize to the target nucleotide sequence. A primer serves as an initiation point for nucleotide polymerization catalyzed by either DNA polymerase, RNA polymerase or reverse transcriptase.

The term "probe" denotes a defined nucleic acid segment (or nucleotide analog segment, e.g., polynucleotide as defined herein) which can be used to identify a specific polynucleotide sequence present in samples, said nucleic acid segment comprising a nucleotide sequence complementary of the specific polynucleotide sequence to be identified.

The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example. Typically the terms "trait" or "phenotype" are used herein to refer to symptoms of, or susceptibility to a disease, a beneficial response to or side effects related to a treatment. Preferably, said trait can be, without being limited to, cancers, developmental diseases, and neurological diseases.

The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic polymorphism has two forms. Typically the first identified allele is designated as the original allele whereas other alleles are designated as alternative alleles. Diploid organisms may be homozygous or heterozygous for an allelic form.

The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population which are heterozygous at a particular allele. In a biallelic system, the heterozygosity rate is on average equal to $2P_a(1-P_a)$, where P_a is the frequency of the least common allele. In order to be useful in genetic studies, a genetic marker should have an adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention, a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker consists of determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

The term "haplotype" refers to a combination of alleles present in an individual or a sample. In the context of the present invention, a haplotype preferably refers to a combination of biallelic marker alleles found in a given individual and which may be associated with a phenotype.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide also gives rise to single nucleotide polymorphisms. In the context of the present invention, "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. Typically, between different individuals, the polymorphic site may be occupied by two different nucleotides.

The term "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a single nucleotide polymorphism having two alleles at a fairly high frequency in the population. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site. Typically, the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (*i.e.* heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (*i.e.* heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker".

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the

polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucl otide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on.

The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another be virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary bases are, generally, A and T (or A and U), or C and G. "Complement" is used herein as a synonym of "complementary polynucleotide", "complementary nucleic acid" and "complementary nucleotide sequence". These terms are applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

Variants and Fragments

1- Polynucleotides

The invention also relates to variants and fragments of the polynucleotides described herein, particularly of a PG-3 gene containing one or more biallelic markers according to the invention.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis

techniques, including those applied to polynucleotides, cells or organisms. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences which are at least 95% identical to a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1 and 2 or to any polynucleotide fragment of at least 12 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1 and 2, and preferably at least 99% identical, more particularly at least 99.5% identical, and most preferably at least 99.8% identical to a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1 and 2, or to any polynucleotide fragment of at least 12 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1 and 2.

Nucleotide changes present in a variant polynucleotide may be silent, which means that they do not alter the amino acids encoded by the polynucleotide. However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

In the context of the present invention, particularly preferred embodiments are those in which the polynucleotides encode polypeptides which retain substantially the same biological function or activity as the mature PG-3 protein, or those in which the polynucleotides encode polypeptides which maintain or increase a particular biological activity, while reducing a second biological activity.

A polynucleotide fragment is a polynucleotide having a sequence that is entirely the same as part but not all of a given nucleotide sequence, preferably the nucleotide sequence of a PG-3 gene, and variants thereof. The fragment can be a portion of an intron or an exon of a PG-3 gene. It can also be a portion of the regulatory regions of PG-3. Preferably, such fragments comprise at least one of the biallelic markers A1 to A80 or the complements thereto or a biallelic marker in linkage disequilibrium with one or more of the biallelic markers A1 to A80.

Such fragments may be "free-standing", *i.e.* not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region. Indeed, several of these fragments may be present within a single larger polynucleotide.

Optionally, such fragments may comprise, consist of, or consist essentially of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 nucleotides in length. A set of preferred fragments contain at least one of the biallelic markers A1 to A80 of the PG-3 gene which are described herein or the complements thereto.

2- Polypeptides

The invention also relates to variants, fragments, analogs and derivatives of the polypeptides described herein, including mutated PG-3 proteins.

The variant may be 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated PG-3 is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mutated PG-3, such as a leader or secretory sequence or a sequence which is employed for purification of the mutated PG-3 or a preprotein sequence. Such variants are deemed to be within the scope of those skilled in the art.

A polypeptide fragment is a polypeptide having a sequence that is entirely the same as part but not all of a given polypeptide sequence, preferably a polypeptide encoded by a PG-3 gene and variants thereof.

In the case of an amino acid substitution in the amino acid sequence of a polypeptide according to the invention, one or several amino acids can be replaced by "equivalent" amino acids. The expression "equivalent" amino acid is used herein to designate any amino acid that may be substituted for one of the amino acids having similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. Generally, the following groups of amino acids represent equivalent changes: (1) Ala, Pro, Gly, Glu, Asp, Gln, Asn, Ser, Thr; (2) Cys, Ser, Tyr, Thr; (3) Val, Ile, Leu, Met, Ala, Phe; (4) Lys, Arg, His; (5) Phe, Tyr, Trp, His.

A specific embodiment of a modified PG-3 peptide molecule of interest according to the present invention, includes, but is not limited to, a peptide molecule which is resistant to proteolysis, a peptide in which the -CONH- peptide bond is modified and replaced by a (CH2NH) reduced bond, a (NHCO) retro inverso bond, a (CH2-O) methylene-oxy bond, a (CH2-S) thiomethylene bond, a (CH2CH2) carba bond, a (CO-CH2) cetomethylene bond, a (CHOH-CH2) hydroxyethylene bond), a (N-N) bound, a E-alcene bond or also a -CH=CH- bond. The invention also encompasses a human PG-3 polypeptide or a fragment or a variant thereof in which at least one peptide bond has been modified as described above.

Such fragments may be "free-standing", *i.e.* not part of or fused to other polypeptides, or they may be included within a single larger polypeptide of which they form a part or region. However, several fragments may be included within a single larger polypeptide.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which are from about 5, 6, 7, 8, 9 or 10 to 15, 10 to 20, 15 to 40, or 30 to 55 amino

acids long. Preferred are those fragments containing at least one amino acid mutation in the PG-3 protein.

Identity Between Nucleic Acids Or P lypeptides

The terms "percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988; Altschul *et al.*, 1990; Thompson *et al.*, 1994; Higgins *et al.*, 1996; Altschul *et al.*, 1993). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990; Altschul *et al.*, 1990, 1993, 1997). In particular, five specific BLAST programs are used to perform the following task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

(3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

(5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (*i.e.*, aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet *et al.*, 1992; Henikoff and Henikoff, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978). The BLAST programs

evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990). The BLAST programs may be used with the default parameters which are implemented in the absence of further instructions from the user. Alternatively, the BLAST programs may be used with parameters specified by the user.

Stringent Hybridization Conditions

By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1X SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2X SSC and 0.1% SDS, or 0.5X SSC and 0.1% SDS, or 0.1X SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and are cited in Sambrook *et al.*, 1989; and Ausubel *et al.*, 1989. These hybridization conditions are suitable for a nucleic acid molecule of about 20 nucleotides in length. There is no need to say that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art. The suitable hybridization conditions may for example be adapted according to the teachings disclosed in Hames and Higgins (1985) or in Sambrook *et al.*(1989).

GENOMIC SEQUENCES OF THE PG-3 GENE

The present invention concerns the genomic sequence of PG-3. The present invention encompasses the PG-3 gene, or PG-3 genomic sequences consisting of, consisting essentially of, or comprising the sequence of SEQ ID No 1, sequences complementary thereto, as well as fragments and variants thereof. These polynucleotides may be purified, isolated, or recombinant.

The invention also encompasses a purified, isolated, or recombinant polynucleotide comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with the nucleotide sequence of SEQ ID No 1 or a complementary sequence thereto or a fragment thereof. The nucleotide differences with regard to the nucleotide sequence of SEQ ID No 1 may be generally randomly distributed throughout the entire nucleic acid. Nevertheless, preferred nucleic

acids are those wherein the nucleotide differences as regards to the nucleotide sequence of SEQ ID No 1 are predominantly located outside the coding sequences contained in the exons. These nucleic acids, as well as their fragments and variants, may be used as oligonucleotide primers or probes in order to detect the presence of a copy of the PG-3 gene in a test sample, or alternatively in order to amplify a target nucleotide sequence within the PG-3 sequences.

Another object of the invention relates to a purified, isolated, or recombinant nucleic acid that hybridizes with the nucleotide sequence of SEQ ID No 1 or a complementary sequence thereto or a variant thereof, under the stringent hybridization conditions as defined above.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825. Additional preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-10000, 10001-20000, 20001-30000, 30001-40000, 40001-50000, 50001-60000, 60001-70000, 70001-80000, 80001-90000, 90001-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-159000, 159001-160000, 160001-170000, 170001-180000, 180001-190000, 190001-200000, 200001-210000, 210001-220000, 220001-230000, 230001-240825. It should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section.

The PG-3 genomic nucleic acid comprises 14 exons. The exon positions in SEQ ID No 1 are detailed below in Table A.

Table A

Exon	Position in SEQ ID No 1		Intron	Position in SEQ ID No 1	
	Beginning	End		Beginning	End
A	2001	2079	A-B	2080	4626
B	4627	4718	B-C	4719	10114
C	10115	10233	C-D	10234	26809
D	26810	26897	D-E	26898	31356
E	31357	31471	E-F	31472	34260
F	34261	34404	F-S	34405	37376
S	37377	37466	S-T	37467	39703
T	39704	40858	T-G	40859	50435
G	50436	50545	G-H	50546	72880
H	72881	72918	H-I	72919	75988
I	75989	76151	I-J	76152	95110

J	95111	95188	J-K	95189	216014
K	216015	216252	K-L	216253	237525
L	237526	238825			

Thus, the invention embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the 14 exons of the PG-3 gene, or a sequence complementary thereto. The invention also relates to purified, isolated, or recombinant nucleic acids comprising a combination of at least two exons of the PG-3 gene, wherein the polynucleotides are arranged within the nucleic acid, from the 5'-end to the 3'-end of said nucleic acid, in the same order as in SEQ ID No 1.

Intron A-B refers to the nucleotide sequence located between Exon A and Exon B, and so on. The position of the introns is detailed in Table A. The intron J-K is large. Indeed, it is 120 kb in length and comprises the whole angiopoietine gene.

Thus, the invention embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the 13 introns of the PG-3 gene, or a sequence complementary thereto.

While this section is entitled "Genomic Sequences of PG-3," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of PG-3 on either side or between two or more such genomic sequences.

PG-3 CDNA SEQUENCES

The expression of the PG-3 gene has been shown to lead to the production of at least one mRNA species which nucleic acid sequence is set forth in SEQ ID No 2. Three cDNAs have been independently cloned. They all have the same size but exhibit strong polymorphism between each other and between each cDNA and the genomic seqeunce. These polymorphisms are indicated in the appended sequence listing by the use of the feature "variation" in SEQ ID No 2.

Another object of the invention is a purified, isolated, or recombinant nucleic acid comprising the nucleotide sequence of SEQ ID No 2, complementary sequences thereto, as well as allelic variants, and fragments thereof. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant PG-3 cDNAs consisting of, consisting essentially of, or comprising the sequence of SEQ ID No 2. Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof. Additional preferred embodiments of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the

following nucleotide positions of SEQ ID No 2: 1-500, 501-1000, 1001-1500, 1501-2000, 2001-2500, 2501-3000, 3001-3500, 3501-3809.

The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 80, 85, 90, or 95% nucleotide identity with a polynucleotide of SEQ ID No 2, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide of SEQ ID No 2, or a sequence complementary thereto or a biologically active fragment thereof.

Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide of SEQ ID No 2, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

The cDNA of SEQ ID No 2 includes a 5'-UTR region starting from the nucleotide at position 1 and ending at the nucleotide in position 57 of SEQ ID No 2. The cDNA of SEQ ID No 2 includes a 3'-UTR region starting from the nucleotide at position 2566 and ending at the nucleotide at position 3809 of SEQ ID No 2. The polyadenylation signal starts from the nucleotide at position 3795 and ends at the nucleotide in position 3800 of SEQ ID No 2.

Consequently, the invention concerns a purified, isolated, or recombinant nucleic acid comprising a nucleotide sequence of the 5'UTR of the PG-3 cDNA, a sequence complementary thereto, or an allelic variant thereof. The invention also concerns a purified, isolated, or recombinant nucleic acid comprising a nucleotide sequence of the 3'UTR of the PG-3 cDNA, a sequence complementary thereto, or an allelic variant thereof.

While this section is entitled "PG-3 cDNA Sequences," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the PG-3 sequences on either side or between two or more such PG-3 sequences.

CODING REGIONS

The PG-3 open reading frame is contained in the corresponding mRNA of SEQ ID No 2. More precisely, the effective PG-3 coding sequence (CDS) includes the region between nucleotide position 58 (first nucleotide of the ATG codon) and nucleotide position 2565 (end nucleotide of the TGA codon) of SEQ ID No 2.

The present invention also embodies isolated, purified, and recombinant polynucleotides which encode a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3. Preferably, the present invention also embodies isolated, purified, and recombinant polynucleotides which encode a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, wherein wherein said contiguous span comprises at least 1, 2, 3, 5, or

10 of the following amino acid positions of SEQ ID No 3: 1-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-835.

The above disclosed polynucleotide that contains the coding sequence of the PG-3 gene may be expressed in a desired host cell or a desired host organism, when this polynucleotide is placed under the control of suitable expression signals. The expression signals may be either the expression signals contained in the regulatory regions in the PG-3 gene of the invention or in contrast the signals may be exogenous regulatory nucleic sequences. Such a polynucleotide, when placed under the suitable expression signals, may also be inserted in a vector for its expression and/or amplification.

REGULATORY SEQUENCES OF PG-3

As mentioned, the genomic sequence of the PG-3 gene contains regulatory sequences both in the non-transcribed 5'-flanking region and in the non-transcribed 3'-flanking region that border the PG-3 coding region containing the 14 exons of this gene.

The 5' regulatory region of the PG-3 gene is localized between the nucleotide in position 1 and the nucleotide in position 2000 of the nucleotide sequence of SEQ ID No 1. The 3' regulatory region of the PG-3 gene is localized between nucleotide position 238826 and nucleotide position 240825 of SEQ ID No 1.

Polynucleotides derived from the 5' and 3' regulatory regions are useful in order to detect the presence of at least a copy of a nucleotide sequence of SEQ ID No 1 or a fragment thereof in a test sample.

The promoter activity of the 5' regulatory regions contained in PG-3 can be assessed as described below.

In order to identify the relevant biologically active polynucleotide fragments or variants of SEQ ID No 1, one of skill in the art will refer to the book of Sambrook *et al.*(1989) which describes the use of a recombinant vector carrying a marker gene (*i.e.* beta galactosidase, chloramphenicol acetyl transferase, etc.) the expression of which will be detected when placed under the control of a biologically active polynucleotide fragments or variants of SEQ ID No 1. Genomic sequences located upstream of the first exon of the PG-3 gene are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, pβgal-Basic, pβgal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech, or pGL2-basic or pGL3-basic promoterless luciferase reporter gene vector from Promega. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, luciferase, β galactosidase, or green fluorescent protein. The sequences upstream the PG-3 coding region are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the

insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for increasing transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Promoter sequences within the upstream genomic DNA may be further defined by constructing nested 5' and/or 3' deletions in the upstream DNA using conventional techniques such as Exonuclease III or appropriate restriction endonuclease digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity, such as described, for example, by Coles *et al.*(1998). In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into cloning sites in promoter reporter vectors. This type of assay is well-known to those skilled in the art and is described in WO 97/17359, US Patent No. 5,374,544; EP 582 796; US Patent No. 5,698,389; US 5,643,746; US Patent No. 5,502,176; and US Patent 5,266,488.

The strength and the specificity of the promoter of the PG-3 gene can be assessed through the expression levels of a detectable polynucleotide operably linked to the PG-3 promoter in different types of cells and tissues. The detectable polynucleotide may be either a polynucleotide that specifically hybridizes with a predefined oligonucleotide probe, or a polynucleotide encoding a detectable protein, including a PG-3 polypeptide or a fragment or a variant thereof. This type of assay is well-known to those skilled in the art and is described in US Patent No. 5,502,176; and US Patent No. 5,266,488. Some of the methods are discussed in more detail below.

Polynucleotides carrying the regulatory elements located at the 5' end and at the 3' end of the PG-3 coding region may be advantageously used to control the transcriptional and translational activity of an heterologous polynucleotide of interest.

Thus, the present invention also concerns a purified or isolated nucleic acid comprising a polynucleotide which is selected from the group consisting of the 5' and 3' regulatory regions, or a sequence complementary thereto or a biologically active fragment or variant thereof.

The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 80, 85, 90, or 95% nucleotide identity with a polynucleotide selected from the group consisting of the 5' and 3' regulatory regions, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide selected from the group consisting of the 5' and 3' regulatory regions, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide selected from the group consisting of the nucleotide sequences of the 5'- and 3' regulatory regions, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

Preferred fragments of the 5' regulatory region have a length of about 1500 or 1000 nucleotides, preferably of about 500 nucleotides, more preferably about 400 nucleotides, even more preferably 300 nucleotides and most preferably about 200 nucleotides.

Preferred fragments of the 3' regulatory region are at least 50, 100, 150, 200, 300 or 400 bases in length.

"Biologically active" polynucleotide derivatives of SEQ ID No 1 are polynucleotides comprising or alternatively consisting essentially of or consisting of a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide in a recombinant cell host. It could act either as an enhancer or as a repressor.

For the purpose of the invention, a nucleic acid or polynucleotide is "functional" as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and translational regulatory information, and such sequences are "operably linked" to nucleotide sequences which encode the desired polypeptide or the desired polynucleotide.

The regulatory polynucleotides of the invention may be prepared from the nucleotide sequence of SEQ ID No 1 by cleavage using suitable restriction enzymes, as described for example in the book of Sambrook *et al.*(1989). The regulatory polynucleotides may also be prepared by digestion of SEQ ID No 1 by an exonuclease enzyme, such as Bal31 (Wabiko *et al.*, 1986). These regulatory polynucleotides can also be prepared by nucleic acid chemical synthesis, as described elsewhere in the specification.

The regulatory polynucleotides according to the invention may be part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism. The recombinant expression vectors according to the invention are described elsewhere in the specification.

A preferred 5'-regulatory polynucleotide of the invention includes the 5'-untranslated region (5'-UTR) of the PG-3 cDNA, or a biologically active fragment or variant thereof.

A preferred 3'-regulatory polynucleotide of the invention includes the 3'-untranslated region (3'-UTR) of the PG-3 cDNA, or a biologically active fragment or variant thereof.

A further object of the invention relates to a purified or isolated nucleic acid comprising:

a) a nucleic acid comprising a regulatory nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence comprising a polynucleotide of the 5' regulatory region or a complementary sequence thereto; or

(ii) a nucleotide sequence comprising a polynucleotide having at least 80, 85, 90, or 95% of nucleotide identity with the nucleotide sequence of the 5' regulatory region or a complementary sequence thereto; or

(iii) a nucleotide sequence comprising a polynucleotide that hybridizes under stringent hybridization conditions with the nucleotide sequence of the 5' regulatory region or a complementary sequence thereto; or

(iv) a biologically active fragment or variant of the polynucleotides in (i), (ii) and (iii);

b) a polynucleotide encoding a desired polypeptide or a nucleic acid of interest, operably linked to the nucleic acid defined in (a) above;

c) Optionally, a nucleic acid comprising a 3'- regulatory polynucleotide, preferably a 3'- regulatory polynucleotide of the PG-3 gene.

In a specific embodiment of the nucleic acid defined above, said nucleic acid includes the 5'-untranslated region (5'-UTR) of the PG-3 cDNA, or a biologically active fragment or variant thereof.

In a second specific embodiment of the nucleic acid defined above, said nucleic acid includes the 3'-untranslated region (3'-UTR) of the PG-3 cDNA, or a biologically active fragment or variant thereof.

The regulatory polynucleotide of the 5' regulatory region, or its biologically active fragments or variants, is operably linked at the 5'-end of the polynucleotide encoding the desired polypeptide or polynucleotide.

The regulatory polynucleotide of the 3' regulatory region, or its biologically active fragments or variants, is advantageously operably linked at the 3'-end of the polynucleotide encoding the desired polypeptide or polynucleotide.

The desired polypeptide encoded by the above-described nucleic acid may be of various nature or origin, encompassing proteins of prokaryotic or eukaryotic origin. Among the polypeptides which may be expressed under the control of a PG-3 regulatory region are bacterial, fungal or viral antigens. Also encompassed are eukaryotic proteins such as intracellular proteins, like "house keeping" proteins, membrane-bound proteins, like receptors, and secreted proteins like endogenous mediators such as cytokines. The desired polypeptide may be the PG-3 protein, especially the protein of the amino acid sequence of SEQ ID No 3, or a fragment or a variant thereof.

The desired nucleic acids encoded by the above-described polynucleotide, usually an RNA molecule, may be complementary to a desired coding p lynucleotide, for example to the PG-3 coding sequence, and thus useful as an antisense polynucleotide.

Such a polynucleotide may be included in a recombinant expression vector in order to express the desired polypeptide or the desired nucleic acid in host cell or in a host organism. Suitable recombinant vectors that contain a polynucleotide such as described herein are disclosed elsewhere in the specification.

POLYNUCLEOTIDE CONSTRUCTS

The terms "polynucleotide construct" and "recombinant polynucleotide" are used interchangeably herein to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their initial natural environment.

DNA Construct That Enables Temporal And Spatial PG-3 Gene Expression In Recombinant Cell Hosts And In Transgenic Animals.

In order to study the physiological and phenotypic consequences of a lack of synthesis of the PG-3 protein, both at the cell level and at the multi cellular organism level, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of the PG-3 genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as regards to the PG-3 nucleotide sequence of SEQ ID Nos 1 and 2, or a fragment thereof, these base substitutions, deletions or additions being located either in an exon, an intron or a regulatory sequence, but preferably in the 5'-regulatory sequence or in an exon of the PG-3 genomic sequence or within the PG-3 cDNA of SEQ ID No 2. In a preferred embodiment, the PG-3 sequence comprises a biallelic marker of the present invention. In a preferred embodiment, the PG-3 sequence comprises at least one of the biallelic markers A1 to A80.

The present invention embodies recombinant vectors comprising any one of the polynucleotides described in the present invention. More particularly, the polynucleotide constructs according to the present invention can comprise any of the polynucleotides described in the "Genomic Sequences Of The PG3 Gene" section, the "PG-3 cDNA Sequences" section, the "Coding Regions" section, and the "Oligonucleotide Probes And Primers" section.

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E. coli* transposon Tn10 for controlling the PG-3 gene expression, such as described by Gossen *et al.*(1992, 1995) and Furth *et al.*(1994). Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tet*op) that are fused to either a minimal promoter or a 5'-regulatory sequence of the PG-3 gene, said minimal promoter or said PG-3 regulatory sequence being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a PG-3 polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (tTA) or the mutant (rTA) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed

under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention comprises both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the tTA or the rTA repressor.

In a specific embodiment, the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor rTA, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

DNA Constructs Allowing Homologous Recombination: Replacement Vectors

A second preferred DNA construct comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is included within the PG-3 genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence that is included within the PG-3 genomic sequence, and is located on the genome downstream the first PG-3 nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream of the nucleotide sequence (a) or downstream from the nucleotide sequence (c). Preferably, the negative selection marker comprises of the thymidine kinase (*tk*) gene (Thomas *et al.*, 1986), the hygromycine beta gene (Te Riele *et al.*, 1990), the *hprt* gene (Van der Lugt *et al.*, 1991; Reid *et al.*, 1990) or the Diphteria toxin A fragment (*Dt-A*) gene (Nada *et al.*, 1993; Yagi *et al.*1990). Preferably, the positive selection marker is located within a PG-3 exon sequence so as to interrupt the sequence encoding a PG-3 protein. These replacement vectors are described, for example, by Thomas *et al.*(1986; 1987), Mansour *et al.*(1988) and Koller *et al.*(1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a PG-3 regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

DNA Constructs Allowing Homologous Recombination: Cre-LoxP System

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *lox*P site. The *lox*P site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess *et al.*, 1986). The recombination by the Cre enzyme between two *lox*P sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*lox*P system used in combination with a homologous recombination technique has been first described by Gu *et al.*(1993, 1994). Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *lox*P sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the

nucleus of the recombinant cell host. The recombinase enzyme may be provided at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki *et al.*(1995), or by lipofection of the enzyme into the cells, such as described by Baubonis *et al.*(1993); (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant host cell, said promoter being optionally inducible, said vector being introduced in the recombinant cell host, such as described by Gu *et al.*(1993) and Sauer *et al.*(1988); (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu *et al.*(1994).

In a specific embodiment, the vector containing the sequence to be inserted in the PG-3 gene by homologous recombination is constructed in such a way that selectable markers are flanked by *lox*P sites of the same orientation, it is possible, by treatment by the Cre enzyme, to eliminate the selectable markers while leaving the PG-3 sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the Cre-*lox*P system are described by Zou *et al.*(1994).

Thus, a third preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is included in the PG-3 genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *lox*P site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is included in the PG-3 genomic sequence, and is located on the genome downstream of the first PG-3 nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *lox*P site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *lox*P sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two loxP sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the Cre enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific prom ter sequence

and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu *et al.*(1994).

The presence of the Cre enzyme within the genome of the recombinant cell host may result from the breeding of two transgenic animals, the first transgenic animal bearing the PG-3-derived sequence of interest containing the *lox*P sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu *et al.*(1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton *et al.* (1995) and Kanegae *et al.*(1995).

The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a PG-3 genomic sequence or a PG-3 cDNA sequence, and most preferably an altered copy of a PG-3 genomic or cDNA sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination). In a specific embodiment, the DNA constructs described above may be used to introduce a PG-3 genomic sequence or a PG-3 cDNA sequence comprising at least one biallelic marker of the present invention, preferably at least one biallelic marker selected from the group consisting of A1 to A80.

Nuclear Antisense DNA Constructs

Other compositions comprise a vector of the invention comprising an oligonucleotide fragment of the nucleic acid sequence of SEQ ID No 2, preferably a fragment including the start codon of the PG-3 gene, as an antisense tool that inhibits the expression of the corresponding PG-3 gene. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel *et al.*(1995) or those described in PCT Application No WO 95/24223.

Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of the PG-3 mRNA. In one embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of PG-3 that contains either the translation initiation codon ATG or a splicing site. Further preferred antisense polynucleotides according to the invention are complementary of the splicing site of the PG-3 mRNA.

Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II

transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu *et al.*(1994). In a preferred embodiment, these PG-3 antisense polynucleotides also comprise, within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as the structure described by Eckner *et al.*(1991).

Oligonucleotide Probes And Primers

Polynucleotides derived from the PG-3 gene are useful in order to detect the presence of at least a copy of a nucleotide sequence of SEQ ID No 1, or a fragment, complement, or variant thereof in a test sample.

Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825. Additional preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-10000, 10001-20000, 20001-30000, 30001-40000, 40001-50000, 50001-60000, 60001-70000, 70001-80000, 80001-90000, 90001-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-159000, 159001-160000, 160001-170000, 170001-180000, 180001-190000, 190001-200000, 200001-210000, 210001-220000, 220001-230000, 230001-240825.

Another object of the invention is a purified, isolated, or recombinant nucleic acid comprising the nucleotide sequence of SEQ ID No 2, complementary sequences thereto, as well as allelic variants, and fragments thereof. Moreover, preferred probes and primers of the invention include purified, isolated, or recombinant PG-3 cDNAs consisting of, consisting essentially of, or comprising the sequence of SEQ ID No 2. Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof. Additional preferred embodiments of the invention include probes and primers comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 2: 1-500, 501-1000, 1001-1500, 1501-2000, 2001-2500, 2501-3000, 3001-3500, 3501-3809.

Thus, the invention also relates to nucleic acid probes characterized in that they hybridize specifically, under the stringent hybridization conditions defined above, with a nucleic acid selected from the group consisting of the nucleotide sequences 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825 of SEQ ID No 1 or a variant thereof or a sequence complementary thereto. The invention relates to nucleic acid probes characterized in that they hybridize specifically, under the stringent hybridization conditions defined above, with a nucleic acid of SEQ ID No 2 or a variant or a fragment thereof or a sequence complementary thereto.

In one embodiment the invention encompasses isolated, purified, and recombinant polynucleotides consisting of, or consisting essentially of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 30, 35, 40, or 50 nucleotides in length of any one of SEQ ID Nos 1 and 2 and the complement thereof, wherein said span includes a PG-3-related biallelic marker in said sequence; optionally, said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said contiguous span is 18 to 35 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide; optionally, said polynucleotide comprises, consists essentially of, or consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide; optionally, the 3' end of said contiguous span is present at the 3' end of said polynucleotide; and optionally, the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide. In a preferred embodiment, said probes comprises, consists of, or consists essentially of a sequence selected from the following sequences: P1 to P4 and P6 to P80 and the complementary sequences thereto.

In another embodiment the invention encompasses isolated, purified or recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 30, 35, 40, or 50 nucleotides in length of SEQ ID Nos 1 and 2, or the complements thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of a PG-3-related biallelic marker in said sequence; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the

complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said PG-3-related biallelic marker in said sequence; and optionally, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D4, D6 to D80, E1 to E4 and E6 to E80.

In a further embodiment, the invention encompasses isolated, purified, or recombinant polynucleotides comprising, consisting of, or consisting essentially of a sequence selected from the following sequences: B1 to B52 and C1 to C52.

In an additional embodiment, the invention encompasses polynucleotides for use in hybridization assays, sequencing assays, and enzyme-based mismatch detection assays for determining the identity of the nucleotide at a PG-3-related biallelic marker in SEQ ID Nos 1 and 2, as well as polynucleotides for use in amplifying segments of nucleotides comprising a PG-3-related biallelic marker in SEQ ID Nos 1 and 2; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith.

The invention concerns the use of the polynucleotides according to the invention for determining the identity of the nucleotide at a PG-3-related biallelic marker, preferably in hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay and in amplifying segments of nucleotides comprising a PG-3-related biallelic marker.

A probe or a primer according to the invention is between 8 and 1000 nucleotides in length, or is specified to be at least 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 or 1000 nucleotides in length. More particularly, the length of these probes and primers can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 15 to 30 nucleotides. Shorter probes and primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes and primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art. A preferred probe or primer consists of a nucleic acid comprising a polynucleotide selected from the group of the

nucleotide sequences of P1 to P4 and P6 to P80 and the complementary sequence thereto, B1 to B52, C1 to C52, D1 to D4, D6 to D80, E1 to E4 and E6 to E80, for which the respective locations in the sequence listing are provided in Tables 1, 2, and 3.

The formation of stable hybrids depends on the melting temperature (Tm) of the DNA. The Tm depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %.

The primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang *et al.*(1979), the phosphodiester method of Brown *et al.*(1979), the diethylphosphoramidite method of Beaucage *et al.*(1981) and the solid support method described in EP 0 707 592.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating any label known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances (including, ^{32}P, ^{35}S, ^{3}H, ^{125}I), fluorescent dyes (including, 5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in the French patent No. FR-7810975, or by Urdea et al (1988) or Sanchez-Pescador et al (1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea *et al.* in 1991 or in the European patent No. EP 0 225 807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the PG-3 gene or mRNA using other techniques.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The

allowing the amplification of a nucleic acid containing the polymorphic base of one PG-3 biallelic marker are listed in Table 1 of Example 2.

Eight PG-3-related biallelic markers A3, A6, A7, A14, A70, A71, A72 and A80, are located in the exonic regions of the genomic sequence of PG-3 at the following positions: 10228, 39944, 39973, 76060, 216026, 216082, 216218 and 237555 of the SEQ ID No 1. They are located in exons C, T, I, K and L of the PG-3 gene. Their respective positions in the cDNA and protein sequences are given in Table 2.

The invention also relates to a purified and/or isolated nucleotide sequence comprising a polymorphic base of a PG-3-related biallelic marker, preferably of a biallelic marker selected from the group consisting of A1 to A80, and the complements thereof. The sequence is between 8 and 1000 nucleotides in length, and preferably comprises at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 or 1000 contiguous nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID Nos 1 and 2 or a variant thereof or a complementary sequence thereto. These nucleotide sequences comprise the polymorphic base of either allele 1 or allele 2 of the considered biallelic marker. Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of said polynucleotide or at the center of said polynucleotide. Optionally, the 3' end of said contiguous span may be present at the 3' end of said polynucleotide. Optionally, biallelic marker may be present at the 3' end of said polynucleotide. Optionally, said polynucleotide may further comprise a label. Optionally, said polynucleotide can be attached to solid support. In a further embodiment, the polynucleotides defined above can be used alone or in any combination.

The invention also relates to a purified and/or isolated nucleotide sequence comprising a sequence between 8 and 1000 nucleotides in length, and preferably at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 or 1000 contiguous nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID Nos 1 and 2 or a variant thereof or a complementary sequence thereto. Optionally, the 3' end of said polynucleotide may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides upstream of a PG-3-related biallelic marker in said sequence. Optionally, said PG-3-related biallelic marker is selected from the group consisting of A1 to A80; Optionally, the 3' end of said polynucleotide may be located 1 nucleotide upstream of a PG-3-related biallelic marker in said sequence. Optionally, said polynucleotide may further comprise a label. Optionally, said polynucleotide can be attached to solid support. In a further embodiment, the polynucleotides defined above can be used alone or in any combination.

In a preferred embodiment, the sequences comprising a polymorphic base of one of the biallelic markers listed in Table 2 are selected from the group consisting of the nucleotide sequences comprising, consisting essentially of, or consisting of the amplicons listed in Table 1 or a variant thereof or a complementary sequence thereto.

solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Consequently, the invention also relates to a method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 1 and 2, a fragment or a variant thereof and a complementary sequence thereto in a sample, said method comprising the following steps of:

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence included in a nucleic acid selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1 and 2, a fragment or a variant thereof and a complementary sequence thereto and the sample to be assayed; and

b) detecting the hybrid complex formed between the probe and a nucleic acid in the sample.

The invention further concerns a kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1 and 2, a fragment or a variant thereof and a complementary sequence thereto in a sample, said kit comprising:

a) a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence included in a nucleic acid selected form the group consisting of the nucleotide sequences of SEQ ID Nos 1 and 2, a fragment or a variant thereof and a complementary sequence thereto; and

b) optionally, the reagents necessary for performing the hybridization reaction.

In a first preferred embodiment of this detection method and kit, said nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule. In a second preferred embodiment of said method and kit, said nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate. In a third preferred embodiment, the nucleic acid probe or the plurality of nucleic acid probes comprise either a sequence which is selected from the group consisting of the nucleotide sequences of P1 to P4 and P6 to P80 and the complementary sequence thereto, B1 to B52, C1 to C52, D1 to D4, D6 to D80, E1 to E4 and E6 to E80 or a biallelic marker selected from the group consisting of A1 to A80 and the complements thereto.

Olig nucleotide Arrays

A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in the PG-3 gene and may also be used for detecting mutations in the coding or in the non-coding sequences of the PG-3 gene.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively, the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotide makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor *et al.*, 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854; and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

In another embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in the PG-3 gene and preferably in its regulatory region. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations, it is meant, mutations on the PG-3 gene that have been identified according, for example to the technique used by Huang *et al.*(1996) or Samson *et al.*(1996).

Another technique that may be used to detect mutations in the PG-3 gene is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of the PG-3 genomic DNA or cDNA. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence within a sample, measure its amount, and detect differences between the target sequence and the sequence of the PG-3 gene in the sample. In one such design, termed 4L tiled array, a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers, is used. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a "footprint" for the probes flanking a mutation position. This technique was described by Chee *et al.* in 1996.

Consequently, the invention concerns an array of nucleic acid molecules comprising at least one polynucleotide described above as probes and primers. Preferably, the invention concerns an array of nucleic acid comprising at least two polynucleotides described above as probes and primers.

A further object of the invention consists of an array of nucleic acid sequences comprising either at least one of the sequences selected from the group consisting of P1 to P4 and P6 to P80, B1 to B52, C1 to C52, D1 to D4, D6 to D80, E1 to E4 and E6 to E80, the sequences complementary thereto, a fragment thereof of at least 8, 10, 12, 15, 18, or 20 consecutive nucleotides thereof, or at least one sequence comprising a biallelic marker selected from the group consisting of A1 to A80 and the complements thereto.

The invention also pertains to an array of nucleic acid sequences comprising either at least two of the sequences selected from the group consisting of P1 to P4, P6 to P80, B1 to B52, C1 to C52, D1 to D4, D6 to D80, E1 to E4 and E6 to E80, the sequences complementary thereto, a fragment thereof of at least 8 consecutive nucleotides thereof, or at least two sequences comprising a biallelic marker selected from the group consisting of A1 to A80 and the complements thereof.

PG-3 PROTEINS AND POLYPEPTIDE FRAGMENTS

The term "PG-3 polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies PG-3 proteins from humans, including isolated or purified PG-3 proteins consisting, consisting essentially, or comprising the sequence of SEQ ID No 3. More particularly, the present invention concerns allelic variants of the PG-3 protein comprising at least one amino acid selected from the group consisting of an arginine or an isoleucine residue at the

amino acid position 304 of the SEQ ID No 3, a histidine or an aspartic acid residue at the amino acid position 314 of the SEQ ID No 3, a threonine or an asparagine residue at the amino acid position 682 of the SEQ ID No 3, an alanine or a valine residue at the amino acid position 761 of the SEQ ID No 3, and a proline or a serine residue at the amino acid position 828 of the SEQ ID No 3. In adddition, the invention also encompasses polypeptide variants of PG-3 comprising at least one amino acid selected from the group consisting of a methionine or an isoleucine residue at the position 91 of SEQ ID No 3, a valine or an alanine residue at the position 306 of SEQ ID No 3, a proline or a serine residue at the position 413 of SEQ ID No 3, a glycine or an aspartate residue at the position 528 of SEQ ID No 3, a valine or an alanine residue at the position 614 of SEQ ID No 3, a threonine or an asparagine residue at the position 677 of SEQ ID No 3, a valine or an alanine residue at the position 756 of SEQ ID No 3, a valine or an alanine residue at the position 758 of SEQ ID No 3, a lysine or a glutamate residue at the position 809 of SEQ ID No 3, and a cysteine or an arginine residue at the position 821 of SEQ ID No 3.

The present invention includes isolated, purified, or recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3. The present invention also embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the following amino acid positions of SEQ ID No 3: 1-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-835. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the PG-3 protein sequence.

The invention also encompasses purified, isolated, or recombinant polypeptides comprising a sequence having at least 70, 75, 80, 85, 90, 95, 98 or 99% nucleotide identity with the sequence of SEQ ID No 3 or a fragment thereof.

PG-3 proteins are preferably isolated from human or mammalian tissue samples or expressed from human or mammalian genes. The PG-3 polypeptides of the invention can be made using routine expression methods known in the art. The polynucleotide encoding the desired polypeptide, is ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems is used in forming recombinant polypeptides, and a summary of some of the more common systems. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt fractionation, chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

In addition, shorter protein fragments is produced by chemical synthesis. Alternatively the proteins of the invention is extracted from cells or tissues of humans or non-human animals. Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

Any PG-3 cDNA, including SEQ ID No 2, may be used to express PG-3 proteins and polypeptides. The nucleic acid encoding the PG-3 protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The PG-3 insert in the expression vector may comprise the full coding sequence for the PG-3 protein or a portion thereof. For example, the PG-3 derived insert may encode a polypeptide comprising at least 10 consecutive amino acids of the PG-3 protein of SEQ ID No 3, preferably least 10 consecutive amino acids including at least 1, 2, 3, 5 or 10 of the following amino acid positions of SEQ ID No 3: 1-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-835.

The expression vector may be any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, *et al.*, and U.S. Patent No. 5,082,767.

In one embodiment, the entire coding sequence of the PG-3 cDNA through the poly A signal of the cDNA is operably linked to a promoter in the expression vector. Alternatively, if the nucleic acid encoding a portion of the PG-3 protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the PG-3 cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The nucleic acid encoding the PG-3 protein or a portion thereof is obtained by PCR from a bacterial vector containing the PG-3 cDNA of SEQ ID No 3 using oligonucleotide primers complementary to the PG-3 cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5'primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the sequence encoding the PG-3 protein or a portion thereof is positioned properly with respect to the poly A signal. The purified fragment obtained from the

resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

The above procedures may also be used to express a mutant PG-3 protein responsible for a detectable phenotype or a portion thereof.

The expressed protein is purified using conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, a solution containing the expressed PG-3 protein or portion thereof, such as a cell extract, is applied to a column having antibodies against the PG-3 protein or portion thereof attached to the chromatography matrix. The expressed protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

To confirm expression of the PG-3 protein or a portion thereof, the proteins expressed from host cells containing an expression vector containing an insert encoding the PG-3 protein or a portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the PG-3 protein or a portion thereof is being expressed. Generally, the band will have the mobility expected for the PG-3 protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Antibodies capable of specifically recognizing the expressed PG-3 protein or a portion thereof are described below.

If antibody production is not possible, the nucleic acids encoding the PG-3 protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the nucleic acid encoding the PG-3 protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera is β-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites are engineered between the β-globin gene or the nickel binding polypeptide and the PG-3 protein or portion thereof. Thus, the two polypeptides of the chimera is separated from one another by protease digestion.

One useful expression vector for generating β-globin chimeric proteins is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, (1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro Express™ Translation Kit (Stratagene).

ANTIBODIES THAT BIND PG-3 POLYPEPTIDES OF THE INVENTION

Any PG-3 polypeptide or whole protein may be used to generate antibodies capable of specifically binding to an expressed PG-3 protein or fragments thereof as described.

One antibody composition of the invention is capable of specifically binding to the PG-3 protein of SEQ ID No 3. For an antibody composition to specifically bind to the PG-3 protein, it must demonstrate at least a 5%, 10%, 15%, 20%, 25%, 50%, or 100% greater binding affinity for PG-3 protein than for another protein in an ELISA, RIA, or other antibody-based binding assay.

The invention also concerns antibody compositions which are specific for variants of the PG-3 protein, more particularly variants comprising at least one amino acid selected from the group consisting of a methionine or an isoleucine residue at the position 91 of SEQ ID No 3, a valine or an alanine residue at the position 306 of SEQ ID No 3, a proline or a serine residue at the position 413 of SEQ ID No 3, a glycine or an aspartate residue at the position 528 of SEQ ID No 3, a valine or an alanine residue at the position 614 of SEQ ID No 3, a threonine or an asparagine residue at the position 677 of SEQ ID No 3, a valine or an alanine residue at the position 756 of SEQ ID No 3, a valine or an alanine residue at the position 758 of SEQ ID No 3, a lysine or a glutamate residue at the position 809 of SEQ ID No 3, and a cysteine or an arginine residue at the position 821 of SEQ ID No 3. More preferably, the invention encompasses antibody compositions which are specific for an allelic variant of the PG-3 protein, more particularly a variant comprising at least one amino acid selected from the group consisting of an arginine or an isoleucine residue at the amino acid position 304 of SEQ ID No 3, a histidine or an aspartic acid residue at the amino acid position 314 of SEQ ID No 3, a threonine or an asparagine residue at the amino acid position 682 of SEQ ID No 3, an alanine or a valine residue at the amino acid position 761 of SEQ ID No 3, and a proline or a serine residue at the amino acid position 828 of SEQ ID No 3.

In a preferred embodiment, the invention concerns antibody compositions, either polyclonal or monoclonal, capable of selectively binding, or selectively bind to an epitope-containing a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3; preferably, said epitope comprises at least 1, 2, 3, 5 or 10 of the following amino acid positions of SEQ ID No 3: 1-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-835.

The inventi n also concerns a purified or isolated antibody capable of specifically binding to a mutated PG-3 protein or to a fragment or variant thereof comprising an epitope of the mutated PG-3 protein. In another preferred embodiment, the present invention concerns an antibody capable of binding to a polypeptide comprising at least 10 consecutive amino acids of a PG-3 protein and including at least one of the amino acids which can be encoded by the trait causing mutations.

In a preferred embodiment, the invention concerns the use in the manufacture of antibodies of a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3; preferably, said contiguous span comprises at least 1, 2, 3, 5 or 10 of the following amino acid positions of SEQ ID No 3: 1-100, 101-200, 201-300, 301-400, 401-500, 501-600, 601-700, 701-835.

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of PG-3 than the one to which antibody binding is desired, and animals which do not express PG-3 (*i.e.* a PG-3 knock out animal as described herein) are particularly useful for preparing antibodies. PG-3 knock out animals will recognize all or most of the exposed regions of a PG-3 protein as foreign antigens, and therefore produce antibodies with a wider array of PG-3 epitopes. Moreover, smaller polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to any one of the PG-3 proteins. In addition, the humoral immune system of animals which produce a species of PG-3 that resembles the antigenic sequence will preferentially recognize the differences between the animal's native PG-3 species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a technique will be particularly useful in obtaining antibodies that specifically bind to any one of the PG-3 proteins.

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

The antibodies of the invention may be labeled using any one of the radioactive, fluorescent or enzymatic labels known in the art.

Consequently, the invention is also directed to a method for specifically detecting the presence of a PG-3 polypeptide according to the invention in a biological sample, said method comprising the following steps :

a) bringing the biological sample into contact with a polyclonal or monoclonal antibody that specifically binds t a PG-3 polypeptide comprising an amino acid sequence of SEQ ID No 3, or to a peptide fragment or variant thereof; and

b) detecting the antigen-antibody complex formed.

The invention also concerns a diagnostic kit for detecting the presence of a PG-3 polypeptide according to the present invention in a biological sample *in vitro* , wherein said kit comprises:

a) a polyclonal or monoclonal antibody that specifically binds to a PG-3 polypeptide comprising the amino acid sequence of SEQ ID No 3, or to a peptide fragment or variant thereof; optionally the antibody may be labeled; and

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent optionally carrying a label, or being able to be recognized itself by a labeled reagent (particularly in the case when the above-mentioned monoclonal or polyclonal antibody itself is not labeled).

PG-3 -RELATED BIALLELIC MARKERS

Advantages Of The Biallelic Markers Of The Present Invention

The PG-3-related biallelic markers of the present invention offer a number of important advantages over other genetic markers such as RFLP (Restriction fragment length polymorphism) and VNTR (Variable Number of Tandem Repeats) markers.

The first generation of markers were RFLPs, which are variations that modify the length of a restriction fragment. But methods used to identify and to type RFLPs are relatively wasteful of materials, effort, and time. The second generation of genetic markers were VNTRs, which can be categorized as either minisatellites or microsatellites. Minisatellites are tandemly repeated DNA sequences present in units of 5-50 repeats which are distributed along regions of the human chromosomes ranging from 0.1 to 20 kilobases in length. Since they present many possible alleles, their informative content is very high. Minisatellites are scored by performing Southern blots to identify the number of tandem repeats present in a nucleic acid sample from the individual being tested. However, there are only 10^4 potential VNTRs that can be typed by Southern blotting. Moreover, both RFLP and VNTR markers are costly and time-consuming to develop and assay in large numbers.

Single nucleotide polymorphisms (SNPs) or biallelic markers can be used in the same manner as RFLPs and VNTRs but offer several advantages. SNPs are densely spaced in the human genome and represent the most frequent type of variation. An estimated number of more than 10^7 sites are scattered along the $3x10^9$ base pairs of the human genome. Therefore, SNPs occur at a greater frequency and with greater uniformity than RFLP or VNTR markers which means that there is a greater probability that such a marker will be found in close proximity to a genetic locus of interest. SNPs are less variable than VNTR markers but are mutationally more stable.

Also, the different forms of a characterized single nucleotide polymorphism, such as the biallelic markers of the present invention, are often easier to distinguish and can therefore be typed easily on a routine basis. Biallelic markers have single nucleotide based alleles and they have only two common alleles, which allows highly parallel detection and automated scoring. The biallelic

markers of the present invention offer the possibility of rapid, high throughput genotyping of a large number of individuals.

Biallelic markers are densely spaced in the genome, sufficiently informative and can be assayed in large numbers. The combined effects of these advantages make biallelic markers extremely valuable in genetic studies. Biallelic markers can be used in linkage studies in families, in allele sharing methods, in linkage disequilibrium studies in populations, in association studies of case-control populations or of trait positive and trait negative populations. An important aspect of the present invention is that biallelic markers allow association studies to be performed to identify genes involved in complex traits. Association studies examine the frequency of marker alleles in unrelated case- and control-populations and are generally employed in the detection of polygenic or sporadic traits. Association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families (linkage studies). Biallelic markers in different genes can be screened in parallel for direct association with disease or response to a treatment. This multiple gene approach is a powerful tool for a variety of human genetic studies as it provides the necessary statistical power to examine the synergistic effect of multiple genetic factors on a particular phenotype, drug response, sporadic trait, or disease state with a complex genetic etiology.

Candidate Gene Of The Present Invention

Different approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. Genome-wide association studies rely on the screening of genetic markers evenly spaced and covering the entire genome. The candidate gene approach is based on the study of genetic markers specifically located in genes potentially involved in a biological pathway related to the trait of interest. In the present invention, PG-3 is a good candidate gene for cancer. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available. However, it should be noted that all of the biallelic markers disclosed in the instant application can be employed as part of genome-wide association studies or as part of candidate region association studies and such uses are specifically contemplated in the present invention and claims.

PG-3-Related Biallelic Markers And Polynucleotides Related Thereto

The invention also concerns PG-3-related biallelic markers. As used herein the term "PG-3-related biallelic marker" relates to a set of biallelic markers in linkage disequilibrium with the PG-3 gene. The term PG-3-related biallelic marker includes the biallelic markers designated A1 to A80.

A portion of the biallelic markers of the present invention are disclosed in Table 2. Their locations in the PG-3 gene are indicated in Table 2 and also as a single base polymorphism in the features of SEQ ID Nos 1 and 2 listed in the accompanying Sequence Listing. The pairs of primers

The invention further concerns a nucleic acid encoding the PG-3 protein, wherein said nucleic acid comprises a polymorphic base of a biallelic marker selected from the group consisting of A1 to A80 and the complements thereof.

The invention also encompasses the use of any polynucleotide for, or any polynucleotide for use in, determining the identity of one or more nucleotides at a PG-3-related biallelic marker. In addition, the polynucleotides of the invention for use in determining the identity of one or more nucleotides at a PG-3-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination. Optionally, said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said PG-3-related biallelic marker is selected from the group consisting A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said polynucleotide may comprise a sequence disclosed in the present specification; Optionally, said polynucleotide may comprise, consist of, or consist essentially of any polynucleotide described in the present specification; Optionally, said determining may involve a hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay; Optionally, said polynucleotide may be attached to a solid support, array, or addressable array; Optionally, said polynucleotide may be labeled. A preferred polynucleotide may be used in a hybridization assay for determining the identity of the nucleotide at a PG-3-related biallelic marker. Another preferred polynucleotide may be used in a sequencing or microsequencing assay for determining the identity of the nucleotide at a PG-3-related biallelic marker. A third preferred polynucleotide may be used in an enzyme-based mismatch detection assay for determining the identity of the nucleotide at a PG-3-related biallelic marker. A fourth preferred polynucleotide may be used in amplifying a segment of polynucleotides comprising a PG-3-related biallelic marker. Optionally, any of the polynucleotides described above may be attached to a solid support, array, or addressable array; Optionally, said polynucleotide may be labeled.

Additionally, the invention encompasses the use of any polynucleotide for, or any polynucleotide for use in amplifying a segment of nucleotides comprising a PG-3-related biallelic marker. In addition, the polynucleotides of the invention for use in amplifying a segment of nucleotides comprising a PG-3-related biallelic marker encompass polynucleotides with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said PG-3-related biallelic marker is selected from the group consisting of A1

to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said PG-3-related biallelic marker is selected from the group consisting A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said polynucleotide may comprise a sequence disclosed in the present specification; Optionally, said polynucleotide may comprise, consist of, or consist essentially of any polynucleotide described in the present specification; Optionally, said amplifying may involve PCR or LCR. Optionally, said polynucleotide may be attached to a solid support, array, or addressable array. Optionally, said polynucleotide may be labeled.

The primers for amplification or sequencing reaction of a polynucleotide comprising a biallelic marker of the invention may be designed from the disclosed sequences for any method known in the art. A preferred set of primers are fashioned such that the 3' end of the contiguous span of identity with a sequence selected from the group consisting of SEQ ID Nos 1 and 2 or a sequence complementary thereto or a variant thereof is present at the 3' end of the primer. Such a configuration allows the 3' end of the primer to hybridize to a selected nucleic acid sequence and dramatically increases the efficiency of the primer for amplification or sequencing reactions. Allele specific primers may be designed such that a polymorphic base of a biallelic marker is at the 3' end of the contiguous span and the contiguous span is present at the 3' end of the primer. Such allele specific primers tend to selectively prime an amplification or sequencing reaction so long as they are used with a nucleic acid sample that contains one of the two alleles present at a biallelic marker. The 3' end of the primer of the invention may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides upstream of a PG-3-related biallelic marker in said sequence or at any other location which is appropriate for their intended use in sequencing, amplification or the location of novel sequences or markers. Thus, another set of preferred amplification primers comprise an isolated polynucleotide consisting essentially of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 30, 35, 40, or 50 nucleotides in length of a sequence selected from the group consisting of SEQ ID Nos 1 and 2 or a sequence complementary thereto or a variant thereof, wherein the 3' end of said contiguous span is located at the 3'end of said polynucleotide, and wherein the 3'end of said polynucleotide is located upstream of a PG-3-related biallelic marker in said sequence. Preferably, those amplification primers comprise a sequence selected from the group consisting of the sequences B1 to B52 and C1 to C52. Primers with their 3' ends located 1 nucleotide upstream of a biallelic marker of PG-3 have a special utility as microsequencing assays. Preferred microsequencing primers are described in Table 4. Optionally, said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, said PG-3-related biallelic marker is selected from the group

consisting A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, microsequencing primers are selected from the group consisting of the nucleotide sequences of D1 to D4, D6 to D80, E1 to E4 and E6 to E80. More preferred microsequencing primers are selected from the group consisting of the nucleotides sequences of D14, D46, D68, D70, D71, E3, E6, E7, E11, E13, E42, E44, E72 and E75.

The probes of the present invention may be designed from the disclosed sequences for use in any method known in the art, particularly methods for testing if a marker disclosed herein is present in a sample. A preferred set of probes may be designed for use in the hybridization assays of the invention in any manner known in the art such that they selectively bind to one allele of a biallelic marker, but not the other under any particular set of assay conditions. Preferred hybridization probes comprise the polymorphic base of either allele 1 or allele 2 of the relevant biallelic marker. Optionally, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of the hybridization probe or at the center of said probe. In a preferred embodiment, the probes are selected from the group consisting of the sequences P1 to P4 and P6 to P80 and the complementary sequence thereto.

It should be noted that the polynucleotides of the present invention are not limited to having the exact flanking sequences surrounding the polymorphic bases which are enumerated in Sequence Listing. Rather, it will be appreciated that the flanking sequences surrounding the biallelic markers may be lengthened or shortened to any extent compatible with their intended use and the present invention specifically contemplates such sequences. The flanking regions outside of the contiguous span need not be homologous to native flanking sequences which actually occur in human subjects. The addition of any nucleotide sequence which is compatible with the polynucleotide's intended use is specifically contemplated.

Primers and probes may be labeled or immobilized on a solid support as described in the section entitled "Oligonucleotide probes and primers".

The polynucleotides of the invention which are attached to a solid support encompass polynucleotides with any further limitation described in this disclosure, or those following, alone or in any combination: Optionally, said polynucleotides may be attached individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. Optionally, polynucleotides other than those of the invention may attached to the same solid support as polynucleotides of the invention. Optionally, when multiple polynucleotides are attached to a solid support they may be attached at random locations, or in an ordered array. Optionally, said ordered array may be addressable.

The present invention also encompasses diagnostic kits comprising one or more polynucleotides of the invention with a portion or all of the necessary reagents and instructions for genotyping a test subject by determining the identity of a nucleotide at a PG-3-related biallelic marker. The polynucleotides of a kit may optionally be attached to a solid support, or be part of an

array or addressable array of polynucleotides. The kit may provide for the determination of the identity of the nucleotide at a marker position by any method known in the art including, but not limited to, a sequencing assay method, a microsequencing assay method, a hybridization assay method, or an enzyme-based mismatch detection assay method.

METHODS FOR *DE NOVO* IDENTIFICATION OF BIALLELIC MARKERS

Any of a variety of methods can be used to screen a genomic fragment for single nucleotide polymorphisms, including methods such as differential hybridization with oligonucleotide probes, detection of changes in the mobility measured by gel electrophoresis or direct sequencing of the amplified nucleic acid. A preferred method for identifying biallelic markers involves comparative sequencing of genomic DNA fragments from an appropriate number of unrelated individuals.

In a first embodiment, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms. One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions, which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained thereby usually demonstrates a sufficient frequency of its less common allele to be useful in conducting association studies.

In a second embodiment, the DNA samples are not pooled and are therefore amplified and sequenced individually. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes. Preferably, highly relevant gene regions such as promoter regions or exon regions may be screened for biallelic markers. A biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele is less than about 10%. Such a biallelic marker will, however, be sufficiently informative to conduct association studies and it will further be appreciated that including less informative biallelic markers in the genetic analysis studies of the present invention, may, in some cases, allow the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations.

The following is a description of the various parameters of a preferred method used by the inventors for the identification of the biallelic markers of the present invention.

Genomic DNA Samples

The genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background. The number of individuals from whom DNA samples are obtained can vary substantially, but is preferably from about 10 to about 1000, or preferably from about 50 to about 200 individuals. It is usually preferred to collect DNA samples from at least

about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA from biological samples are well known to the skilled technician. Details of a preferred embodiment are provided in Example 1. The person skilled in the art can choose to amplify pooled or unpooled DNA samples.

DNA Amplification

The identification of biallelic markers in a sample of genomic DNA may be facilitated through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well known to those skilled in the art.

Amplification techniques that can be used in the context of the present invention include, but are not limited to, the ligase chain reaction (LCR) described in EP-A- 320 308, WO 9320227 and EP-A-439 182, the polymerase chain reaction (PCR, RT-PCR) and techniques such as the nucleic acid sequence based amplification (NASBA) described in Guatelli J.C., *et al.*(1990) and in Compton J.(1991), Q-beta amplification as described in European Patent Application No 4544610, strand displacement amplification as described in Walker *et al.*(1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461.

LCR and Gap LCR are exponential amplification techniques, both of which utilize DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in th first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes, which can be

ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall *et al.*(1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see White (1992) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195; 4,683,202; and 4,965,188.

The PCR technology is the preferred amplification technique used to identify new biallelic markers. A typical example of a PCR reaction suitable for the purposes of the present invention is provided in Example 2.

One of the aspects of the present invention is a method for the amplification of the human PG-3 gene, particularly of a fragment of the genomic sequence of SEQ ID No 1 or of the cDNA sequence of SEQ ID No 2, or a fragment or a variant thereof in a test sample, preferably using the PCR technology. This method comprises the steps of:

a) contacting a test sample with amplification reaction reagents comprising a pair of amplification primers as described above which are located on either side of the polynucleotide region to be amplified, and

b) optionally, detecting the amplification products.

The invention also concerns a kit for the amplification of a PG-3 gene sequence, particularly of a portion of the genomic sequence of SEQ ID No 1 or of the cDNA sequence of SEQ ID No 2, or a variant thereof in a test sample, wherein said kit comprises:

a) a pair of oligonucleotide primers located on either side of the PG-3 region to be amplified;

b) optionally, the reagents necessary for performing the amplification reaction.

In one embodiment of the above amplification method and kit, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region. In another embodiment of the above amplification method and kit, primers comprise a sequence which is selected from the group consisting of the nucleotide sequences of B1 to B52, C1 to C52, D1 to D4, D6 to D80, E1 to E4, and E6 to E80.

In a first embodiment of the present invention, biallelic markers are identified using genomic sequence information generated by the inventors. Sequenced genomic DNA fragments are used to design primers for the amplification of 500 bp fragments. These 500 bp fragments are amplified from genomic DNA and are scanned for biallelic markers. Primers may be designed using the OSP software (Hillier L. and Green P., 1991). All primers may contain, upstream of the specific target bases, a common oligonucleotide tail that serves as a sequencing primer. Those skilled in the art are familiar with primer extensions, which can be used for these purposes.

Preferred primers, useful for the amplification of genomic sequences encoding the candidate genes, focus on promoters, exons and splice sites of the genes. A biallelic marker presents a higher probability to be a causal mutation if it is located in these functional regions of the gene. Preferred amplification primers of the invention include the nucleotide sequences B1 to B52 and C1 to C52, detailed further in Example 2, Table 1.

Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms

The amplification products generated as described above, are then sequenced using any method known and available to the skilled technician. Methods for sequencing DNA using either the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to those of ordinary skill in the art. Such methods are disclosed in Sambrook *et al.*(1989) for example. Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee *et al.*(1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. The products of the sequencing reactions are run on sequencing gels and the sequences are determined using gel image analysis. The polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. Because each dideoxy terminator is labeled with a different fluorescent molecule, the two peaks corresponding to a biallelic site present distinct colors corresponding to two different nucleotides at the same position on the sequence. However, the presence of two peaks can be an artifact due to background noise. To exclude such an

artifact, the two DNA strands are sequenced and a comparison between the peaks is carried out. In order to confirm that a sequence is polymorphic, the polymorphism is be detected on both strands.

The above procedure permits those amplification products which contain biallelic markers to be identified. The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is approximately 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies. However, more than 90% of the biallelic polymorphisms detected by the pooling method have a frequency for the minor allele higher than 0.25. Therefore, the biallelic markers selected by this method have a frequency of at least 0.1 for the minor allele and less than 0.9 for the major allele. Preferably, the biallelic markers selected by this method have a frequency of at least 0.2 for the minor allele and less than 0.8 for the major allele, more preferably at least 0.3 for the minor allele and less than 0.7 for the major allele. Thus, the biallelic markers preferably have a heterozygosity rate higher than 0.18, more preferably higher than 0.32, still more preferably higher than 0.42.

In another embodiment, biallelic markers are detected by sequencing individual DNA samples. In some embodiments, the frequency of the minor allele of such a biallelic marker may be less than 0.1.

Validation Of The Biallelic Markers Of The Present Invention

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals, so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a sequence. All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with validated biallelic markers.

Evaluation Of The Frequency Of The Biallelic Markers Of The Present Invention

The validated biallelic markers are further evaluated for their usefulness as genetic markers by determining the frequency of the least common allele at the biallelic marker site. The higher the frequency of the less common allele the greater the usefulness of the biallelic marker in association

and interaction studies. The identification of the least common allele is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. The determination of marker frequency by genotyping may be performed using individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group must be large enough to be representative of the population as a whole. Preferably the group contains at least 20 individuals, more preferably the group contains at least 50 individuals, most preferably the group contains at least 100 individuals. Of course the larger the group the greater the accuracy of the frequency determination because of reduced sampling error. A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker." All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with high quality biallelic markers.

METHODS FOR GENOTYPING AN INDIVIDUAL FOR BIALLELIC MARKERS

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which may be performed *in vitro*. Such methods of genotyping comprise determining the identity of a nucleotide at a PG-3 biallelic marker site by any method known in the art. These methods find use in genotyping case-control populations in association studies as well as individuals in the context of detection of alleles of biallelic markers which are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual may be classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed on nucleic acid samples derived from a single individual or pooled DNA samples.

Genotyping can be performed using methods similar to those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

In one embodiment, the invention encompasses methods of genotyping comprising determining the identity of a nucleotide at a PG-3-related biallelic marker or the complement thereof in a biological sample; optionally, the PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof,

or optionally the biallelic markers in linkage disequilibrium therewith; optionally, the biological sample is derived from a single subject; optionally, the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome; optionally, said biological sample is derived from multiple subjects; Optionally, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following, alone or in any combination; Optionally, said method is performed *in vitro*; optionally, the method further comprises amplifying a portion of said sequence comprising the biallelic marker prior to said determining step; Optionally, the amplifyication is performed by PCR, LCR, or replication of a recombinant vector comprising an origin of replication and said fragment in a host cell; optionally, the determination involves a hybridization assay, a sequencing assay, a microsequencing assay, or an enzyme-based mismatch detection assay.

Source of Nucleic Acids for genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described above. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

Amplification Of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention. Amplification of DNA may be achieved by any method known in the art. Amplification techniques are described above in the section entitled, "DNA amplification."

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as further described below.

The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention.

In some embodiments, the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. Preferred amplification primers are listed in Example 2. It will be appreciated that the primers listed are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention are also of use.

The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention, amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers may be any sequence which allow the specific amplification of any DNA fragment carrying the markers. Amplification primers may be labeled or immobilized on a solid support as described in the section "Oligonucleotide probes and primers".

Methods of Genotyping DNA samples for Biallelic Markers

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove simple for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita *et al.*(1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield *et al.*(1991), White *et al.*(1992), Grompe *et al.*(1989 and 1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, enzyme-based mismatch detection assay, or hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing" is generally used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

1) Sequencing Assays

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as

described above. DNA sequ ncing methods are described in th section entitled "Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms".

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

2) Microsequencing Assays

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in Example 4.

Different approaches can be used for the labeling and detection of ddNTPs. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (1997) and Chen *et al.*(1997). In this method, amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein–labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff and Smirnov, 1997).

Microsequencing may be achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is conducted as a heterogeneous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase

extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator regent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner, oligonucleotides or templates may be attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, 1994) or linked to fluorescein (Livak and Hainer, 1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju *et al.*, 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren *et al.*(1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen *et al.*(1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described below.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. Preferred microsequencing primers include the nucleotide sequences D1 to D4 and D6 to D80 and E1 to E4 and E6 to E80. It will be appreciated that the microsequencing primers listed in Example 4 are merely exemplary and that any primer having a 3' end immediately adjacent to the polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for any biallelic marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Example 4, or fragments comprising at least 8, 12, 15, 20, 25, 30, 40, or 50 consecutive nucleotides thereof, to the extent that such lengths are consistent with

the primer described, and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a nucleotide at a biallelic marker site.

3) Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of polymerases and ligases. Polymerization reactions place particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above in the section entitled "Amplification Of DNA Fragments Comprising Biallelic Markers".

Allele Specific Amplification Primers

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy whereby one of the alleles is amplified without amplification of the other allele. For allele specific amplification, at least one member of the pair of primers is sufficiently complementary with a region of a PG-3 gene comprising the polymorphic base of a biallelic marker of the present invention to hybridize therewith and to initiate the amplification. Such primers are able to discriminate between the two alleles of a biallelic marker.

This is accomplished by placing the polymorphic base at the 3' end of one of the amplification primers. Because the extension progresses from the 3'end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Determining the precise location of the mismatch and the corresponding assay conditions are well within the ordinary skill in the art.

Ligation/Amplification Based Methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting single nucleotide polymorphisms and may be advantageously combined with PCR as described by Nickerson *et al.*(1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other amplification methods which are particularly suited for the detection of single nucleotide polymorphism include LCR (ligase chain reaction), Gap LCR (GLCR) which are described above in the section entitled "DNA Amplification". LCR uses two pairs of probes to

exponentially amplify a specific target. The sequences of each pair of oligonucleotides are selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit Analysis™ is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

4) Hybridization Assay Methods

A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook *et al.*, 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization c nditions, under which a probe will hybridize only to the exactly complementary target sequence

are well known in the art (Sambrook *et al.*, 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. Although such hybridization can be performed in solution, it is preferred to employ a solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. *et al.*, 1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak *et al.*, 1995). In an alternative homogeneous hybridization based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi *et al.*, 1998).

The polynucleotides provided herein can be used to produce probes which can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes preferably comprise between 8 and 50 nucleotides and are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. A particularly preferred probe is 25 nucleotides in length. Preferably the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes, the biallelic marker is at the center of said polynucleotide. Preferred probes comprise a nucleotide sequence selected from the group consisting of amplicons listed in Table 1 and the sequences

complementary thereto, or a fragment thereof, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. Preferred probes compris a nucleotide sequence selected from the group consisting of P1 to P4 and P6 to P80 and the sequences complementary thereto. In preferred embodiments the polymorphic base(s) are within 5, 4, 3, 2, 1, nucleotides of the center of the said polynucleotide, more preferably at the center of said polynucleotide.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in the section entitled "Oligonucleotide Probes and Primers". The probes can be non-extendable as described in the section entitled "Oligonucleotide Probes and Primers".

By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample. High-Throughput parallel hybridization in array format is specifically encompassed within "hybridization assays" and is described below.

5) Hybridization To Addressable Arrays Of Oligonucleotides

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (e.g., the chip) at selected positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia *et al.*, 1996; Shoemaker *et al.*, 1996; Kozal *et al.*, 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP 785280, describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular, the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of

biallelic markers. For example, a detection block may be tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To obtain probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186.

Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences about 15 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of amplicons listed in Table 1 and the sequences complementary thereto, or a fragment thereof, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. In preferred embodiments the polymorphic base is within 5, 4, 3, 2, 1, nucleotides of the center of the said polynucleotide, more preferably at the center of said polynucleotide. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more of these polynucleotides of the invention. Solid supports and polynucleotides of the present invention attached to solid supports are further described in the section entitled "Oligonucleotide Probes And Primers".

6) Integrated Systems

Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts.

For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

METHODS OF GENETIC ANALYSIS USING THE BIALLELIC MARKERS OF THE PRESENT INVENTION

Different methods are available for the genetic analysis of complex traits (see Lander and Schork, 1994). The search for disease-susceptibility genes is conducted using two main methods: the linkage approach in which evidence is sought for cosegregation between a locus and a putative trait locus using family studies, and the association approach in which evidence is sought for a statistically significant association between an allele and a trait or a trait causing allele (Khoury *et al.*, 1993). In general, the biallelic markers of the present invention find use in any method known in the art to demonstrate a statistically significant correlation between a genotype and a phenotype. The biallelic markers may be used in parametric and non-parametric linkage analysis methods. Preferably, the biallelic markers of the present invention are used to identify genes associated with detectable traits using association studies, an approach which does not require the use of affected families and which permits the identification of genes associated with complex and sporadic traits.

The genetic analysis using the biallelic markers of the present invention may be conducted on any scale. The whole set of biallelic markers of the present invention or any subset of biallelic markers of the present invention corresponding to the candidate gene may be used. Further, any set of genetic markers including a biallelic marker of the present invention may be used. A set of biallelic polymorphisms that could be used as genetic markers in combination with the biallelic markers of the present invention has been described in WO 98/20165. As mentioned above, it should be noted that the biallelic markers of the present invention may be included in any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

Linkage Analysis

Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. Thus, the aim of linkage analysis is to detect marker loci that show cosegregation with a trait of interest in pedigrees.

PARAMETRIC METHODS

When data are available from successive generations there is the opportunity to study the degree of linkage between pairs of loci. Estimates of the recombination fraction enable loci to be ordered and placed onto a genetic map. With loci that are genetic markers, a genetic map can be established, and then the strength of linkage between markers and traits can be calculated and used to indicate the relative positions of markers and genes affecting those traits (Weir, 1996). The classical method for linkage analysis is the logarithm of odds (lod) score method (see Morton, 1955; Ott, 1991). Calculation of lod scores requires specification of th mode of inheritance for the

disease (parametric method). Generally, the length of the candidate region identified using linkage analysis is between 2 and 20Mb. Once a candidate region is identified as described above, analysis of recombinant individuals using additional markers allows further delineation of the candidate region. Linkage analysis studies have generally relied on the use of a maximum of 5,000 microsatellite markers, thus limiting the maximum theoretical attainable resolution of linkage analysis to about 600 kb on average.

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (*i.e.*, the ratio between the number of trait positive carriers of allele *a* and the total number of *a* carriers in the population). However, parametric linkage analysis suffers from a variety of drawbacks. First, it is limited by its reliance on the choice of a genetic model suitable for each studied trait. Furthermore, as already mentioned, the resolution attainable using linkage analysis is limited, and complementary studies are required to refine the analysis of the typical 2Mb to 20Mb regions initially identified through linkage analysis. In addition, parametric linkage analysis approaches have proven difficult when applied to complex genetic traits, such as those due to the combined action of multiple genes and/or environmental factors. It is very difficult to model these factors adequately in a lod score analysis. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (1996).

NON-PARAMETRIC METHODS

The advantage of the so-called non-parametric methods for linkage analysis is that they do not require specification of the mode of inheritance for the disease, they tend to be more useful for the analysis of complex traits. In non-parametric methods, one tries to prove that the inheritance pattern of a chromosomal region is not consistent with random Mendelian segregation by showing that affected relatives inherit identical copies of the region more often than expected by chance. Affected relatives should show excess "allele sharing" even in the presence of incomplete penetrance and polygenic inheritance. In non-parametric linkage analysis the degree of agreement at a marker locus in two individuals can be measured either by the number of alleles identical by state (IBS) or by the number of alleles identical by descent (IBD). Affected sib pair analysis is a well-known special case and is the simplest form of these methods.

The biallelic markers of the present invention may be used in both parametric and non-parametric linkage analysis. Preferably biallelic markers may be used in non-parametric methods which allow the mapping of genes involved in complex traits. The biallelic markers of the present invention may be used in both IBD- and IBS- methods to map genes affecting a complex trait. In such studies, taking advantage of the high density of biallelic markers, several adjacent biallelic marker loci may be pooled to achieve the efficiency attained by multi-allelic markers (Zhao *et al.*, 1998).

Populati n Associati n Studies

The present invention comprises methods for detecting an association between the PG-3 gene and a detectable trait using the biallelic markers of the present invention. In one embodiment the present invention comprises methods to detect an association between a biallelic marker allele or a biallelic marker haplotype and a trait. Further, the invention comprises methods to identify a trait causing allele in linkage disequilibrium with any biallelic marker allele of the present invention.

As described above, alternative approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. In a preferred embodiment, the biallelic markers of the present invention are used to perform candidate gene association studies. The candidate gene analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available. Further, the biallelic markers of the present invention may be incorporated in any map of genetic markers of the human genome in order to perform genome-wide association studies. Methods to generate a high-density map of biallelic markers has been described in US Provisional Patent application serial number 60/082,614. The biallelic markers of the present invention may further be incorporated in any map of a specific candidate region of the genome (a specific chromosome or a specific chromosomal segment for example).

As mentioned above, association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families. Association studies are extremely valuable as they permit the analysis of sporadic or multifactor traits. Moreover, association studies represent a powerful method for fine-scale mapping enabling much finer mapping of trait causing alleles than linkage studies. Studies based on pedigrees often only narrow the location of the trait causing allele. Association studies using the biallelic markers of the present invention can therefore be used to refine the location of a trait causing allele in a candidate region identified by Linkage Analysis methods. Moreover, once a chromosome segment of interest has been identified, the presence of a candidate gene such as a candidate gene of the present invention, in the region of interest can provide a shortcut to the identification of the trait causing allele. Biallelic markers of the present invention can be used to demonstrate that a candidate gene is associated with a trait. Such uses are specifically contemplated in the present invention.

Determining The Frequency Of A Biallelic Marker Allele Or Of A Biallelic Marker Haplotype In A Population

Association studies explore the relationships among frequencies for sets of alleles between loci.

DETERMINING THE FREQUENCY OF AN ALLELE IN A POPULATION

Allelic frequencies of the biallelic markers in a populations can be determined using one of the methods described above under the heading "Methods for genotyping an individual for biallelic markers", or any genotyping procedure suitable for this intended purpose. Genotyping pooled samples or individual samples can determine the frequency of a biallelic marker allele in a population. One way to reduce the number of genotypings required is to use pooled samples. A drawback in using pooled samples is in terms of accuracy and reproducibility for determining accurate DNA concentrations in setting up the pools. Genotyping individual samples provides higher sensitivity, reproducibility and accuracy and; is the preferred method used in the present invention. Preferably, each individual is genotyped separately and simple gene counting is applied to determine the frequency of an allele of a biallelic marker or of a genotype in a given population.

The invention also relates to methods of estimating the frequency of an allele in a population comprising: a) genotyping individuals from said population for said biallelic marker according to the method of the present invention; b) determining the proportional representation of said biallelic marker in said population. In addition, the methods of estimating the frequency of an allele in a population of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination; optionally, the PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic marker is one of the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, the determination of the frequency of a biallelic marker allele in a population may be accomplished by determining the identity of the nucleotides for both copies of said biallelic marker present in the genome of each individual in said population and calculating the proportional representation of said nucleotide at said PG-3-related biallelic marker for the population; Optionally, the determination of the proportional representation may be accomplished by performing a genotyping method of the invention on a pooled biological sample derived from a representative number of individuals, or each individual, in said population, and calculating the proportional amount of said nucleotide compared with the total.

DETERMINING THE FREQUENCY OF A HAPLOTYPE IN A POPULATION

The gametic phase of haplotypes is unknown when diploid individuals are heterozygous at more than one locus. Using genealogical information in families gametic phase can sometimes be inferred (Perlin *et al.*, 1994). When no genealogical information is available different strategies may be used. One possibility is that the multiple-site heterozygous diploids can be eliminated from the analysis, keeping only the homozygotes and the single-site heterozygote individuals, but this approach might lead to a possible bias in the sample composition and the underestimation of low-

frequency haplotypes. Another possibility is that single chromosomes can be studied independently, for example, by asymmetric PCR amplification (see Newton et al, 1989; Wu *et al.*, 1989) or by isolation of single chromosome by limit dilution followed by PCR amplification (see Ruano *et al.*, 1990). Further, a sample may be haplotyped for sufficiently close biallelic markers by double PCR amplification of specific alleles (Sarkar, G. and Sommer S. S., 1991). These approaches are not entirely satisfying either because of their technical complexity, the additional cost they entail, their lack of generalization at a large scale, or the possible biases they introduce. To overcome these difficulties, an algorithm to infer the phase of PCR-amplified DNA genotypes introduced by Clark, A.G.(1990) may be used. Briefly, the principle is to start filling a preliminary list of haplotypes present in the sample by examining unambiguous individuals, that is, the complete homozygotes and the single-site heterozygotes. Then other individuals in the same sample are screened for the possible occurrence of previously recognized haplotypes. For each positive identification, the complementary haplotype is added to the list of recognized haplotypes, until the phase information for all individuals is either resolved or identified as unresolved. This method assigns a single haplotype to each multiheterozygous individual, whereas several haplotypes are possible when there are more than one heterozygous site. Alternatively, one can use methods estimating haplotype frequencies in a population without assigning haplotypes to each individual. Preferably, a method based on an expectation-maximization (EM) algorithm (Dempster *et al.*, 1977) leading to maximum-likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions (random mating) is used (see Excoffier L. and Slatkin M., 1995). The EM algorithm is a generalized iterative maximum-likelihood approach to estimation that is useful when data are ambiguous and/or incomplete. The EM algorithm is used to resolve heterozygotes into haplotypes. Haplotype estimations are further described below under the heading "Statistical Methods." Any other method known in the art to determine or to estimate the frequency of a haplotype in a population may be used.

The invention also encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping at least one PG-3-related biallelic marker according to a method of the invention for each individual in said population; b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. In addition, the methods of estimating the frequency of a haplotype of the invention encompass methods with any further limitation described in this disclosure, or those following, alone or in any combination: optionally, said PG-3-related biallelic marker is selected from the gr up consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the

group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said haplotype determination method is performed by asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark algorithm, or an expectation-maximization algorithm.

Linkage Disequilibrium Analysis

Linkage disequilibrium is the non-random association of alleles at two or more loci and represents a powerful tool for mapping genes involved in disease traits (see Ajioka R.S. *et al.*, 1997). Biallelic markers, because they are densely spaced in the human genome and can be genotyped in greater numbers than other types of genetic markers (such as RFLP or VNTR markers), are particularly useful in genetic analysis based on linkage disequilibrium.

When a disease mutation is first introduced into a population (by a new mutation or the immigration of a mutation carrier), it necessarily resides on a single chromosome and thus on a single "background" or "ancestral" haplotype of linked markers. Consequently, there is complete disequilibrium between these markers and the disease mutation: one finds the disease mutation only in the presence of a specific set of marker alleles. Through subsequent generations recombination events occur between the disease mutation and these marker polymorphisms, and the disequilibrium gradually dissipates. The pace of this dissipation is a function of the recombination frequency, so the markers closest to the disease gene will manifest higher levels of disequilibrium than those that are further away. When not broken up by recombination, "ancestral" haplotypes and linkage disequilibrium between marker alleles at different loci can be tracked not only through pedigrees but also through populations. Linkage disequilibrium is usually seen as an association between one specific allele at one locus and another specific allele at a second locus.

The pattern or curve of disequilibrium between disease and marker loci is expected to exhibit a maximum that occurs at the disease locus. Consequently, the amount of linkage disequilibrium between a disease allele and closely linked genetic markers may yield valuable information regarding the location of the disease gene. For fine-scale mapping of a disease locus, it is useful to have some knowledge of the patterns of linkage disequilibrium that exist between markers in the studied region. As mentioned above the mapping resolution achieved through the analysis of linkage disequilibrium is much higher than that of linkage studies. The high density of biallelic markers combined with linkage disequilibrium analysis provides powerful tools for fine-scale mapping. Different methods to calculate linkage disequilibrium are described below under the heading "Statistical Methods".

Population-Based Case-Control Studies Of Trait-Marker Associations

As mentioned above, the occurrence of pairs of specific alleles at different loci on the same chromosome is not random and the deviation from random is called linkage disequilibrium.

Association studies focus on population frequencies and rely on the phenomenon of linkage disequilibrium. If a specific allele in a given gene is directly involved in causing a particular trait, its frequency will be statistically increased in an affected (trait positive) population, when compared to the frequency in a trait negative population or in a random control population. As a consequence of the existence of linkage disequilibrium, the frequency of all other alleles present in the haplotype carrying the trait-causing allele will also be increased in trait positive individuals compared to trait negative individuals or random controls. Therefore, association between the trait and any allele (specifically a biallelic marker allele) in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular region. Case-control populations can be genotyped for biallelic markers to identify associations that narrowly locate a trait causing allele. As any marker in linkage disequilibrium with one given marker associated with a trait will be associated with the trait. Linkage disequilibrium allows the relative frequencies in case-control populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analyzed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles. Association studies compare the frequency of marker alleles in unrelated case-control populations, and represent powerful tools for the dissection of complex traits.

CASE-CONTROL POPULATIONS (INCLUSION CRITERIA)

Population-based association studies do not concern familial inheritance but compare the prevalence of a particular genetic marker, or a set of markers, in case-control populations. They are case-control studies based on comparison of unrelated case (affected or trait positive) individuals and unrelated control (unaffected, trait negative or random) individuals. Preferably the control group is composed of unaffected or trait negative individuals. Further, the control group is ethnically matched to the case population. Moreover, the control group is preferably matched to the case-population for the main known confusion factor for the trait under study (for example age-matched for an age-dependent trait). Ideally, individuals in the two samples are paired in such a way that they are expected to differ only in their disease status. The terms "trait positive population", "case population" and "affected population" are used interchangeably herein.

An important step in the dissection of complex traits using association studies is the choice of case-control populations (see Lander and Schork, 1994). A major step in the choice of case-control populations is the clinical definition of a given trait or phenotype. Any genetic trait may be analyzed by the association method proposed here by carefully selecting the individuals to be included in the trait positive and trait negative phenotypic groups. Four criteria are often useful: clinical phenotype, age at onset, family history and severity. The selection procedure for continuous or quantitative traits (such as blood pressure for example) involves selecting individuals at opposite ends of the phenotype distribution of the trait under study, so as to include in these trait positive and trait negative populations individuals with non-overlapping phenotypes. Preferably, case-control populations consist of phenotypically homogeneous populations. Trait positive and

More particularly, the present invention relates to expression vectors which include nucleic acids encoding a PG-3 protein, preferably the PG-3 protein of the amino acid sequence of SEQ ID No 3 or variants or fragments thereof.

The invention also pertains to a recombinant expression vector useful for the expression of the PG-3 coding sequence, wherein said vector comprises a nucleic acid of SEQ ID No 2.

Recombinant vectors comprising a nucleic acid containing a PG-3-related biallelic marker are also part of the invention. In a preferred embodiment, said biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof.

Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

The present invention also encompasses primary, secondary, and immortalized homologously recombinant host cells of vertebrate origin, preferably mammalian origin and particularly human origin, that have been engineered to: a) insert exogenous (heterologous) polynucleotides into the endogenous chromosomal DNA of a targeted gene, b) delete endogenous chromosomal DNA, and/or c) replace endogenous chromosomal DNA with exogenous polynucleotides. Insertions, deletions, and/or replacements of polynucleotide sequences may be to the coding sequences of the targeted gene and/or to regulatory regions, such as promoter and enhancer sequences, operably associated with the targeted gene.

The present invention further relates to a method of making a homologously recombinant host cell in vitro or in vivo, wherein the expression of a targeted gene not normally expressed in the cell is altered. Preferably the alteration causes expression of the targeted gene under normal growth conditions or under conditions suitable for producing the polypeptide encoded by the targeted gene. The method comprises the steps of: (a) transfecting the cell in vitro or in vivo with a polynucleotide construct, the polynucleotide construct comprising; (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; and (b) maintaining the transfected cell in vitro or in vivo under conditions appropriate for homologous recombination.

The present invention further relates to a method of altering the expression of a targeted gene in a cell in vitro or in vivo wherein the gene is not normally expressed in the cell, comprising the steps of: (a) transfecting the cell in vitro or in vivo with a a polynucleotide construct, the a polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; and (b) maintaining the transfected cell in vitro or in vivo under conditions appropriate for homologous recombination, thereby producing a homologously recombinant cell; and (c) maintaining the homologously recombinant cell in vitro or in vivo under conditions appropriate for expression of the gene.

trait negative populations consist of phenotypically uniform populations of individuals representing each between 1 and 98%, preferably between 1 and 80%, more preferably between 1 and 50%, and more preferably between 1 and 30%, most preferably between 1 and 20% of the total population under study, and preferably selected among individuals exhibiting non-overlapping phenotypes. The clearer the difference between the two trait phenotypes, the greater the probability of detecting an association with biallelic markers. The selection of those drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible detection of marked differences at the genetic level, provided that the sample sizes of the populations under study are significant enough.

In preferred embodiments, a first group of between 50 and 300 trait positive individuals, preferably about 100 individuals, are recruited according to their phenotypes. A similar number of control individuals are included in such studies.

ASSOCIATION ANALYSIS

The invention also comprises methods of detecting an association between a genotype and a phenotype, comprising the steps of: a) determining the frequency of at least one PG-3-related biallelic marker in a trait positive population according to a genotyping method of the invention; b) determining the frequency of said PG-3-related biallelic marker in a control population according to a genotyping method of the invention; and c) determining whether a statistically significant association exists between said genotype and said phenotype. In addition, the methods of detecting an association between a genotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said control population may be a trait negative population, or a random population; Optionally, each of said genotyping steps a) and b) may be performed on a pooled biological sample derived from each of said populations; Optionally, each of said genotyping of steps a) and b) is performed separately on biological samples derived from each individual in said population or a subsample thereof; Optionally, said trait is cancer susceptibility.

The general strategy to perform association studies using biallelic markers derived from a region carrying a candidat gene is to scan two groups of individuals (case-control populations) in order to measure and statistically compare the allele frequencies of the biallelic markers of the present invention in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analyzed biallelic markers, one can assume that: either the associated allele is directly responsible for causing the trait (*i.e.* the associated allele is the trait causing allele), or more likely the associated allele is in linkage disequilibrium with the trait causing allele. The specific characteristics of the associated allele with respect to the candidate gene function usually give further insight into the relationship between the associated allele and the trait (causal or in linkage disequilibrium). If the evidence indicates that the associated allele within the candidate gene is most probably not the trait causing allele but is in linkage disequilibrium with the real trait causing allele, then the trait causing allele can be found by sequencing the vicinity of the associated marker, and performing further association studies with the polymorphisms that are revealed in an iterative manner.

Association studies are usually run in two successive steps. In a first phase, the frequencies of a reduced number of biallelic markers from the candidate gene are determined in the trait positive and control populations. In a second phase of the analysis, the position of the genetic loci responsible for the given trait is further refined using a higher density of markers from the relevant region. However, if the candidate gene under study is relatively small in length, as is the case for PG-3, a single phase may be sufficient to establish significant associations.

HAPLOTYPE ANALYSIS

As described above, when a chromosome carrying a disease allele first appears in a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a set of linked markers: the ancestral haplotype. This haplotype can be tracked through populations and its statistical association with a given trait can be analyzed. Complementing single point (allelic) association studies with multi-point association studies also called haplotype studies increases the statistical power of association studies. Thus, a haplotype association study allows one to define the frequency and the type of the ancestral carrier haplotype. A haplotype analysis is important in that it increases the statistical power of an analysis involving individual markers.

In a first stage of a haplotype frequency analysis, the frequency of the possible haplotypes based on various combinations of the identified biallelic markers of the invention is determined. The haplotype frequency is then compared for distinct populations of trait positive and control individuals. The number of trait positive individuals, which should be, subjected to this analysis to obtain statistically significant results usually ranges between 30 and 300, with a preferred number of individuals ranging between 50 and 150. The same considerations apply to the number of unaffected individuals (or random control) used in the study. The results of this first analysis provide haplotype frequencies in case-control populations, for each evaluated haplotype frequency a p-value and an odd ratio are calculated. If a statistically significant association is found the relative

risk for an individual carrying the given haplotype of being affected with the trait under study can be approximated.

An additional embodiment of the present invention encompasses methods of detecting an association between a haplotype and a phenotype, comprising the steps of: a) estimating the frequency of at least one haplotype in a trait positive population, according to a method of the invention for estimating the frequency of a haplotype; b) estimating the frequency of said haplotype in a control population, according to a method of the invention for estimating the frequency of a haplotype; and c) determining whether a statistically significant association exists between said haplotype and said phenotype. In addition, the methods of detecting an association between a haplotype and a phenotype of the invention encompass methods with any further limitation described in this disclosure, or those following: optionally, said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; Optionally, said control population is a trait negative population, or a random population. Optionally, said method comprises the additional steps of determining the phenotype in said trait positive and said control populations prior to step c); optionally, said trait is cancer susceptibility.

INTERACTION ANALYSIS

The biallelic markers of the present invention may also be used to identify patterns of biallelic markers associated with detectable traits resulting from polygenic interactions. The analysis of genetic interaction between alleles at unlinked loci requires individual genotyping using the techniques described herein. The analysis of allelic interaction among a selected set of biallelic markers with an appropriate level of statistical significance can be considered as a haplotype analysis. Interaction analysis consists in stratifying the case-control populations with respect to a given haplotype for the first loci and performing a haplotype analysis with the second loci with each subpopulation.

Statistical methods used in association studies are further described below.

Testing For Linkage In The Presence Of Association

The biallelic markers of the present invention may further be used in TDT (transmission/disequilibrium test). TDT tests for both linkage and association and is not affected by population stratification. TDT requires data for affected individuals and their parents or data from unaffected sibs instead of from parents (see Spielmann S. *et al.*, 1993; Schaid D.J. *et al.*, 1996, Spielmann S. and Ewens W.J., 1998). Such combined tests generally reduce the false – positive errors produced by separate analyses.

STATISTICAL METHODS

In general, any method known in the art to test whether a trait and a genotype show a statistically significant correlation may be used.

1) Methods In Linkage Analysis

Statistical methods and computer programs useful for linkage analysis are well-known to those skilled in the art (see Terwilliger J.D. and Ott J., 1994; Ott J., 1991).

2) Methods To Estimate Haplotype Frequencies In A Population

As described above, when genotypes are scored, it is often not possible to distinguish heterozygotes so that haplotype frequencies cannot be easily inferred. When the gametic phase is not known, haplotype frequencies can be estimated from the multilocus genotypic data. Any method known to person skilled in the art can be used to estimate haplotype frequencies (see Lange K., 1997; Weir, B.S., 1996) Preferably, maximum-likelihood haplotype frequencies are computed using an Expectation- Maximization (EM) algorithm (see Dempster *et al.*, 1977; Excoffier L. and Slatkin M., 1995). This procedure is an iterative process aiming at obtaining maximum-likelihood estimates of haplotype frequencies from multi-locus genotype data when the gametic phase is unknown. Haplotype estimations are usually performed by applying the EM algorithm using for example the EM-HAPLO program (Hawley M. E. *et al.*, 1994) or the Arlequin program (Schneider *et al.*, 1997). The EM algorithm is a generalized iterative maximum likelihood approach to estimation and is briefly described below.

Please note that in the present section, "Methods To Estimate Haplotype Frequencies In A Population, ", phenotypes will refer to multi-locus genotypes with unknown haplotypic phase. Genotypes will refer to mutli-locus genotypes with known haplotypic phase.

Suppose one has a sample of N unrelated individuals typed for K markers. The data observed are the unknown-phase K-locus phenotypes that can be categorized with F different phenotypes. Further, suppose that we have H possible haplotypes (in the case of K biallelic markers, we have for the maximum number of possible haplotypes $H=2^K$).

For phenotype j with cj possible genotypes, we have:

$$P_j = \sum_{i=1}^{c_j} P(genotype(i)) = \sum_{i=1}^{c_j} P(h_k, h_l). \qquad \text{Equation 1}$$

Here, P_j is the probability of the j^{th} phenotype, and $P(h_k,h_l)$ is the probability of the i^{th} genotype composed of haplotypes h_k and h_l. Under random mating (*i.e.* Hardy-Weinberg Equilibrium), $P(h_kh_l)$ is expressed as:

$$P(h_k, h_l) = P(h_k)^2 \text{ for } h_k = h_l, \text{ and}$$

$$P(h_k, h_l) = 2P(h_k)P(h_l) \text{ for } h_k \neq h_l. \qquad \text{Equation 2}$$

The E-M algorithm is composed of the following steps: First, the genotype frequencies are estimated from a set of initial values of haplotype frequencies. These haplotype frequencies are

denoted $P_1^{(0)}, P_2^{(0)}, P_3^{(0)},\ldots, P_H^{(0)}$. The initial values for the haplotype frequencies may be obtained from a random number generator or in some other way well known in the art. This step is referred to the Expectation step. The next step in the method, called the Maximization step, consists of using the estimates for the genotype frequencies to re-calculate the haplotype frequencies. The first iteration haplotype frequency estimates are denoted by $P_1^{(1)}, P_2^{(1)}, P_3^{(1)},\ldots, P_H^{(1)}$. In general, the Expectation step at the s^{th} iteration consists of calculating the probability of placing each phenotype into the different possible genotypes based on the haplotype frequencies of the previous iteration:

$$P(h_k,h_l)^{(s)} = \frac{n_j}{N}\left[\frac{P_j(h_k,h_l)^{(s)}}{P_j}\right], \qquad \text{Equation 3}$$

where n_j is the number of individuals with the j^{th} phenotype and $P_j(h_k,h_l)^{(s)}$ is the probability of genotype $h_k h_l$ in phenotype j. In the Maximization step, which is equivalent to the gene-counting method (Smith, 1957), the haplotype frequencies are re-estimated based on the genotype estimates:

$$P_t^{(s+1)} = \frac{1}{2}\sum_{j=1}^{F}\sum_{i=1}^{c_j}\delta_{it}P_j(h_k,h_l)^{(s)}. \qquad \text{Equation 4}$$

Here, δ_{it} is an indicator variable which counts the number of occurrences that haplotype t is present in i^{th} genotype; it takes on values 0, 1, and 2.

The E-M iterations cease when the following criterion has been reached. Using Maximum Likelihood Estimation (MLE) theory, one assumes that the phenotypes j are distributed multinomially. At each iteration s, one can compute the likelihood function L. Convergence is achieved when the difference of the log-likehood between two consecutive iterations is less than some small number, preferably 10^{-7}.

3) Methods To Calculate Linkage Disequilibrium Between Markers

A number of methods can be used to calculate linkage disequilibrium between any two genetic positions, in practice linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population.

Linkage disequilibrium between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention (M_i, M_j) having alleles (a_i/b_i) at marker M_i and alleles (a_j/b_j) at marker M_j can be calculated for every allele combination (a_i,a_j; a_i,b_j; b_i,a_j and b_i,b_j), according to the Piazza formula:

$\Delta_{aiaj} = \sqrt{\theta 4} - \sqrt{(\theta 4 + \theta 3)(\theta 4 + \theta 2)}$, where:

$\theta 4$= - - = frequency of genotypes not having allele a_i at M_i and not having allele a_j at M_j

$\theta 3$= - + = frequency of genotypes not having allele a_i at M_i and having allele a_j at M_j

$\theta 2$= + - = frequency of genotypes having allele a_i at M_i and not having allele a_j at M_j

Linkage disequilibrium (LD) between pairs of biallelic markers (M_i, M_j) can also be calculated for every allele combination (ai,aj; ai,bj; b_i,a_j and b_i,b_j), according to the maximum-

likelihood estimate (MLE) for delta (the composite genotypic disequilibrium coefficient), as described by Weir (Weir B. S., 1996). The MLE for the composite linkage disequilibrium is:

$D_{aiaj} = (2n_1 + n_2 + n_3 + n_4/2)/N - 2(pr(a_i).\ pr(a_j))$

Where n_1 = Σ phenotype (a_i/a_i, a_j/a_j), n_2 = Σ phenotype (a_i/a_i, a_j/b_j), n_3= Σ phenotype (a_i/b_i, a_j/a_j), n4= Σ phenotype (a_i/b_i, a_j/b_j) and N is the number of individuals in the sample.

This formula allows linkage disequilibrium between alleles to be estimated when only genotype, and not haplotype, data are available.

Another means of calculating the linkage disequilibrium between markers is as follows. For a couple of biallelic markers, M_i (a_i/b_i) and M_j (a_j/b_j), fitting the Hardy-Weinberg equilibrium, one can estimate the four possible haplotype frequencies in a given population according to the approach described above.

The estimation of gametic disequilibrium between *ai* and *aj* is simply:

$$D_{aiaj} = pr(haplotype(a_i, a_j)) - pr(a_i).pr(a_j).$$

Where $pr(a_i)$ is the probability of allele a_i and pr(a_j) is the probability of allele a_j and where $pr(haplotype\ (a_i, a_j))$ is estimated as in Equation 3 above.

For a couple of biallelic marker only one measure of disequilibrium is necessary to describe the association between M_i and M_j.

Then a normalized value of the above is calculated as follows:

$D'_{aiaj} = D_{aiaj} / \max(-pr(a_i).\ pr(a_j), -pr(b_i).\ pr(b_j))$ with $D_{aiaj}<0$

$D'_{aiaj} = D_{aiaj} / \max(pr(b_i).\ pr(a_j), pr(a_i).\ pr(b_j))$ with $D_{aiaj}>0$

The skilled person will readily appreciate that other linkage disequilibrium calculation methods can be used.

Linkage disequilibrium among a set of biallelic markers having an adequate heterozygosity rate can be determined by genotyping between 50 and 1000 unrelated individuals, preferably between 75 and 200, more preferably around 100.

4) Testing For Association

Methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case an allele at a biallelic marker or a haplotype made up of such alleles, may be determined by any statistical test known in the art and with any accepted threshold of statistical significance being required. The application of particular methods and thresholds of significance are well with in the skill of the ordinary practitioner of the art.

Testing for association is performed by determining the frequency of a biallelic marker allele in case and control populations and comparing these frequencies with a statistical test to determine if their is a statistically significant difference in frequency which would indicate a correlation between the trait and the biallelic marker allele under study. Similarly, a haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in case and control populations, and comparing these frequencies with a statistical

test to determine if their is a statistically significant correlation between the haplotype and the phenotype (trait) under study. Any statistical tool useful to test for a statistically significant association between a genotype and a phenotype may be used. Preferably the statistical test employed is a chi-square test with one degree of freedom. A P-value is calculated (the P-value is the probability that a statistic as large or larger than the observed one would occur by chance).

STATISTICAL SIGNIFICANCE

In preferred embodiments, significance for diagnosis purposes, either as a positive basis for further diagnostic tests or as a preliminary starting point for early preventive therapy, the p value related to a biallelic marker association is preferably about 1×10^{-2} or less, more preferably about 1×10^{-4} or less, for a single biallelic marker analysis and about 1×10^{-3} or less, still more preferably 1×10^{-6} or less and most preferably of about 1×10^{-8} or less, for a haplotype analysis involving two or more markers. These values are believed to be applicable to any association studies involving single or multiple marker combinations.

The skilled person can use the range of values set forth above as a starting point in order to carry out association studies with biallelic markers of the present invention. In doing so, significant associations between the biallelic markers of the present invention and a trait can be revealed and used for diagnosis and drug screening purposes.

PHENOTYPIC PERMUTATION

In order to confirm the statistical significance of the first stage haplotype analysis described above, it might be suitable to perform further analyses in which genotyping data from case-control individuals are pooled and randomized with respect to the trait phenotype. Each individual genotyping data is randomly allocated to two groups, which contain the same number of individuals as the case-control populations used to compile the data obtained in the first stage. A second stage haplotype analysis is preferably run on these artificial groups, preferably for the markers included in the haplotype of the first stage analysis showing the highest relative risk coefficient. This experiment is reiterated preferably at least between 100 and 10000 times. The repeated iterations allow the determination of the probability to obtain the tested haplotype by chance.

ASSESSMENT OF STATISTICAL ASSOCIATION

To address the problem of false positives similar analysis may be performed with the same case-control populations in random genomic regions. Results in random regions and the candidate region are compared as described in a co-pending US Provisional Patent Application entitled "Methods, Software And Apparati For Identifying Genomic Regions Harboring A Gene Associated With A Detectable Trait," U.S. Serial Number 60/107,986, filed November 10, 1998, and a second U.S. Provisional Patent Application also entitled "Methods, Software And Apparati For Identifying Genomic Regions Harboring A Gene Associated With A Detectable Trait," U.S. Serial Number 60/140,785, filed June 23, 1999.

5) Evaluation Of Risk Factors

The association between a risk factor (in genetic epidemiology the risk factor is the presence or the absence of a certain allele or haplotype at marker loci) and a disease is measured by the odds ratio (OR) and by the relative risk (RR). If $P(R^+)$ is the probability of developing the disease for individuals with R and $P(R^-)$ is the probability for individuals without the risk factor, then the relative risk is simply the ratio of the two probabilities, that is:

$RR= P(R^+)/P(R^-)$

In case-control studies, direct measures of the relative risk cannot be obtained because of the sampling design. However, the odds ratio allows a good approximation of the relative risk for low-incidence diseases and can be calculated:

$$OR = \left[\frac{F^+}{1-F^+}\right] \Big/ \left[\frac{F^-}{(1-F^-)}\right]$$

$OR= (F^+/(1-F^+))/(F^-/(1-F^-))$

F^+ is the frequency of the exposure to the risk factor in cases and F^- is the frequency of the exposure to the risk factor in controls. F^+ and F^- are calculated using the allelic or haplotype frequencies of the study and further depend on the underlying genetic model (dominant, recessive, additive...).

One can further estimate the attributable risk (AR) which describes the proportion of individuals in a population exhibiting a trait due to a given risk factor. This measure is important in quantifying the role of a specific factor in disease etiology and in terms of the public health impact of a risk factor. The public health relevance of this measure lies in estimating the proportion of cases of disease in the population that could be prevented if the exposure of interest were absent. AR is determined as follows:

$AR = P_E (RR-1) / (P_E (RR-1)+1)$

AR is the risk attributable to a biallelic marker allele or a biallelic marker haplotype. P_E is the frequency of exposure to an allele or a haplotype within the population at large; and RR is the relative risk which, is approximated with the odds ratio when the trait under study has a relatively low incidence in the general population.

IDENTIFICATION OF BIALLELIC MARKERS IN LINKAGE DISEQUILIBRIUM WITH THE BIALLELIC MARKERS OF THE INVENTION

Once a first biallelic marker has been identified in a genomic region of interest, the practitioner of ordinary skill in the art, using the teachings of the present invention, can easily identify additional biallelic markers in linkage disequilibrium with this first marker. As mentioned before, any marker in linkage disequilibrium with a first marker associated with a trait will be associated with the trait. Therefore, once an association has been demonstrated between a given biallelic marker and a trait, the discovery of additional biallelic markers associated with this trait is of great interest in order to increase the density of biallelic markers in this particular region. The

causal gene or mutation will be found in the vicinity of the marker or set of markers showing the highest correlation with the trait.

Identification of additional markers in linkage disequilibrium with a given marker involves: (a) amplifying a genomic fragment comprising a first biallelic marker from a plurality of individuals; (b) identifying of second biallelic markers in the genomic region harboring said first biallelic marker; (c) conducting a linkage disequilibrium analysis between said first biallelic marker and second biallelic markers; and (d) selecting said second biallelic markers as being in linkage disequilibrium with said first marker. Subcombinations comprising steps (b) and (c) are also contemplated.

Methods to identify biallelic markers and to conduct linkage disequilibrium analysis are described herein and can be carried out by the skilled person without undue experimentation. The present invention then also concerns biallelic markers which are in linkage disequilibrium with the biallelic markers A1 to A80 and which are expected to present similar characteristics in terms of their respective association with a given trait.

IDENTIFICATION OF FUNCTIONAL MUTATIONS

Mutations in the PG-3 gene which are responsible for a detectable phenotype or trait may be identified by comparing the sequences of the PG-3 gene from trait positive and control individuals. Once a positive association is confirmed with a biallelic marker of the present invention, the identified locus can be scanned for mutations. In a preferred embodiment, functional regions such as exons and splice sites, promoters and other regulatory regions of the PG-3 gene are scanned for mutations. In a preferred embodiment the sequence of the PG-3 gene is compared in trait positive and control individuals. Preferably, trait positive individuals carry the haplotype shown to be associated with the trait and trait negative individuals do not carry the haplotype or allele associated with the trait. The detectable trait or phenotype may comprise a variety of manifestations of altered PG-3 function.

The mutation detection procedure is essentially similar to that used for biallelic marker identification. The method used to detect such mutations generally comprises the following steps:

- amplification of a region of the PG-3 gene comprising a biallelic marker or a group of biallelic markers associated with the trait from DNA samples of trait positive patients and trait-negative controls;
- sequencing of the amplified region;
- comparison of DNA sequences from trait positive and control individuals;
- determination of mutations specific to trait-positive patients.

In one embodiment, said biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof. It is preferred that candidate polymorphisms be then verified by screening a larger population of cases and controls by means of any genotyping procedure such as those described herein, preferably using a microsequencing technique in an individual test format.

Polymorphisms are considered as candidate mutations when present in cases and controls at frequencies compatible with the expected association results. Polymorphisms are considered as candidate "trait-causing" mutations when they exhibit a statistically significant correlation with the detectable phenotype.

RECOMBINANT VECTORS

The term "vector" is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism.

The present invention encompasses a family of recombinant vectors that comprise a regulatory polynucleotide derived from the PG-3 genomic sequence, and/or a coding polynucleotide from either the PG-3 genomic sequence or the cDNA sequence.

Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences, coding sequences and polynucleotide constructs, as well as any PG-3 primer or probe as defined above. More particularly, the recombinant vectors of the present invention can comprise any of the polynucleotides described in the "Genomic Sequences Of The PG3 Gene" section, the "PG-3 cDNA Sequences" section, the "Coding Regions" section, the "Polynucleotide constructs" section, and the "Oligonucleotide Probes And Primers" section.

In a first preferred embodiment, a recombinant vector of the invention is used to amplify the inserted polynucleotide derived from a PG-3 genomic sequence of SEQ ID No 1 or a PG-3 cDNA, for example the cDNA of SEQ ID No 2 in a suitable cell host, this polynucleotide being amplified at every time that the recombinant vector replicates.

A second preferred embodiment of the recombinant vectors according to the invention comprises expression vectors comprising either a regulatory polynucleotide or a coding nucleic acid of the invention, or both. Within certain embodiments, expression vectors are employed to express the PG-3 polypeptide, which can then be purified and, for example be used in ligand screening assays or as an immunogen in order to raise specific antibodies directed against the PG-3 protein. In other embodiments, the expression vectors are used for constructing transgenic animals and also for gene therapy. Expression requires that appropriate signals are provided in the vectors, said signals including various regulatory elements, such as enhancers/promoters from both viral and mammalian sources that drive expression of the genes of interest in host cells. Dominant drug selection markers for establishing permanent, stable cell clones expressing the products are generally included in the expression vectors of the invention, as they are elements that link expression of the drug selection markers to expression of the polypeptide.

The present invention further relates to a method of making a polypeptide of the present invention by altering the expression of a targeted endogenous gene in a cell in vitro or in vivo wherein the gene is not normally expressed in the cell, comprising the steps of: a) transfecting the cell in vitro with a a polynucleotide construct, the a polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; (b) maintaining the transfected cell in vitro or in vivo under conditions appropriate for homologous recombination, thereby producing a homologously recombinant cell; and c) maintaining the homologously recombinant cell in vitro or in vivo under conditions appropriate for expression of the gene thereby making the polypeptide.

The present invention further relates to a polynucleotide construct which alters the expression of a targeted gene in a cell type in which the gene is not normally expressed. This occurs when the a polynucleotide construct is inserted into the chromosomal DNA of the target cell, wherein the a polynucleotide construct comprises: a) a targeting sequence; b) a regulatory sequence and/or coding sequence; and c) an unpaired splice-donor site, if necessary. Further included are a polynucleotide constructs, as described above, wherein the construct further comprises a polynucleotide which encodes a polypeptide and is in-frame with the targeted endogenous gene after homologous recombination with chromosomal DNA.

The compositions may be produced, and methods performed, by techniques known in the art, such as those described in U.S. Patent Nos: 6,054,288; 6,048,729; 6,048,724; 6,048,524; 5,994,127; 5,968,502; 5,965,125; 5,869,239; 5,817,789; 5,783,385; 5,733,761; 5,641,670; 5,580,734 ; International Publication Nos:WO96/29411, WO 94/12650; and scientific articles including Koller *et al.*,1989.

1. General features of the expression vectors of the invention

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may consist of a chromosomal, non-chromosomal, semi-synthetic and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of:

(1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp in length that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, said structural or coding sequence being operably linked to the regulatory elements described in (1); and

(3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, when a recombinant

protein is expressed without a leader or transport sequence, it may include a N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of the translated protein into the periplasmic space or the extracellular medium. In a specific embodiment wherein the vector is adapted for transfecting and expressing desired sequences in mammalian host cells, preferred vectors will comprise an origin of replication in the desired host, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation signal, splice donor and acceptor sites, transcriptional termination sequences, and 5'-flanking non-transcribed sequences. DNA sequences derived from the SV40 viral genome, for example SV40 origin, early promoter, enhancer, splice and polyadenylation signals may be used to provide the required non-transcribed genetic elements.

The *in vivo* expression of a PG-3 polypeptide of SEQ ID No 3 or fragments or variants thereof may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism or to the production of a biologically inactive PG-3 protein.

Consequently, the present invention also deals with recombinant expression vectors mainly designed for the *in vivo* production of the PG-3 polypeptide of SEQ ID No 3 or fragments or variants thereof by the introduction of the appropriate genetic material in the organism of the patient to be treated. This genetic material may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue.

2. Regulatory Elements

PROMOTERS

The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account the cell host in which the heterologous gene has to be expressed. The particular promoter employed to control the expression of a nucleic acid sequence of interest is not believed to be important, so long as it is capable of directing the expression of the nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell, such as, for example, a human or a viral promoter.

A suitable promoter may be heterologous with respect to the nucleic acid for which it controls the expression or alternatively can be endogenous to the native polynucleotide containing the coding sequence to be expressed. Additionally, the promoter is generally heterologous with

respect to the recombinant vector sequences within which the construct promoter/coding sequence has been inserted.

Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors.

Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the gpt, lambda PR, PL and trp promoters (EP 0036776), the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith *et al.*, 1983; O'Reilly *et al.*, 1992), the lambda PR promoter or also the trc promoter.

Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

The choice of a promoter is well within the ability of a person skilled in the field of genetic egineering. For example, one may refer to the book of Sambrook *et al.*(1989) or also to the procedures described by Fuller *et al.*(1996).

OTHER REGULATORY ELEMENTS

Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed such as human growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

3. Selectable Markers

Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression construct. The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicin or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria, this latter marker being a negative selection marker.

4. Preferred Vectors.

BACTERIAL VECTORS

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of other suitable vectors are known to those of skill in the art, and commercially available, such as the following bacterial vectors: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A

(Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); pQE-30 (QIAexpress).

BACTERIOPHAGE VECTORS

The P1 bacteriophage vector may contain large inserts ranging from about 80 to about 100 kb.

The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg (1992, 1994). Recombinant P1 clones comprising PG-3 nucleotide sequences may be designed for inserting large polynucleotides of more than 40 kb (Linton *et al.*, 1993). To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick *et al.*(1994). Briefly, *E. coli* (preferably strain NS3529) harboring the P1 plasmid are grown overnight in a suitable broth medium containing 25 µg/ml of kanamycin. The P1 DNA is prepared from the *E. coli* by alkaline lysis using the Qiagen Plasmid Maxi kit (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

When the goal is to express a P1 clone comprising PG-3 nucleotide sequences in a transgenic animal, typically in transgenic mice, it is desirable to remove vector sequences from the P1 DNA fragment, for example by cleaving the P1 DNA at rare-cutting sites within the P1 polylinker (*Sfi*I, *Not*I or *Sal*I). The P1 insert is then purified from vector sequences on a pulsed-field agarose gel, using methods similar using methods similar to those originally reported for the isolation of DNA from YACs (Schedl *et al.*, 1993a; Peterson *et al.*, 1993). At this stage, the resulting purified insert DNA can be concentrated, if necessary, on a Millipore Ultrafree-MC Filter Unit (Millipore, Bedford, MA, USA – 30,000 molecular weight limit) and then dialyzed against microinjection buffer (10 mM Tris-HCl, pH 7.4; 250 µM EDTA) containing 100 mM NaCl, 30 µM spermine, 70 µM spermidine on a microdyalisis membrane (type VS, 0.025 µM from Millipore). The intactness of the purified P1 DNA insert is assessed by electrophoresis on 1% agarose (Sea Kem GTG; FMC Bio-products) pulse-field gel and staining with ethidium bromide.

BACULOVIRUS VECTORS

A suitable vector for the expression of the PG-3 polypeptide of SEQ ID No 3 or fragments or variants thereof is a baculovirus vector that can be propagated in insect cells and in insect cell lines. A specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector (Pharmingen) that is used to transfect the SF9 c ll line (ATCC N°CRL 1711) which is derived from *Spodoptera frugiperda*.

Other suitable vectors for the expression of th PG-3 polypeptide of SEQ ID No 3 or fragments or variants thereof in a baculovirus expression system include those described by Chai *et al.*(1993), Vlasak *et al.*(1983) and Lenhard *et al.*(1996).

VIRAL VECTORS

In one specific embodiment, the vector is derived from an adenovirus. Preferred adenovirus vectors according to the invention are those described by Feldman and Steg (1996) or Ohno *et al.*(1994). Another preferred recombinant adenovirus according to this specific embodiment of the present invention is the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French patent application N° FR-93.05954).

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery systems of choice for the transfer of exogenous polynucleotides *in vivo* , particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro* gene delivery vehicles of the present invention include retroviruses selected from the group consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include the 4070A and the 1504A viruses, Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-190; PCT Application No WO 94/24298). Particularly preferred Rous Sarcoma Viruses include Bryan high titer (ATCC Nos VR-334, VR-657, VR-726, VR-659 and VR-728). Other preferred retroviral vectors are those described in Roth *et al.*(1996), PCT Application No WO 93/25234, PCT Application No WO 94/ 06920, Roux *et al.*, 1989, Julan *et al.*, 1992 and Neda *et al.*, 1991.

Yet another viral vector system that is contemplated by the invention consists in the adeno-associated virus (AAV). The adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka *et al.*, 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte *et al.*, 1992; Samulski *et al.*, 1989; McLaughlin *et al.*, 1989). One advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

BAC VECTORS

The bacterial artificial chromosome (BAC) cloning system (Shizuya *et al.*, 1992) has been developed to stably maintain large fragments of genomic DNA (100-300 kb) in *E. coli*. A preferred BAC vector consists of pBeloBAC11 vector that has been described by Kim *et al.*(1996). BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into either the *Bam* HI or *Hind*III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites

that can be used to generate end probes by either RNA transcription or PCR methods. After the construction of a BAC library in *E. coli*, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The cloning site is flanked by two *Not* I sites, permitting cloned segments to be excised from the vector by *Not* I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector may be linearized by treatment of the BAC vector with the commercially available enzyme lambda terminase that leads to the cleavage at the unique *cos*N site, but this cleavage method results in a full length BAC clone containing both the insert DNA and the BAC sequences.

5. Delivery Of The Recombinant Vectors

In order to effect expression of the polynucleotides and polynucleotide constructs of the invention, these constructs must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cell lines, or *in vivo* or *ex vivo*, as in the treatment of certain diseases states.

One mechanism is viral infection where the expression construct is encapsulated in an infectious viral particle.

Several non-viral methods for the transfer of polynucleotides into cultured mammalian cells are also contemplated by the present invention, and include, without being limited to, calcium phosphate precipitation (Graham *et al.*, 1973; Chen *et al.*, 1987;), DEAE-dextran (Gopal, 1985), electroporation (Tur-Kaspa *et al.*, 1986; Potter *et al.*, 1984), direct microinjection (Harland *et al.*, 1985), DNA-loaded liposomes (Nicolau *et al.*, 1982; Fraley *et al.*, 1979), and receptor-mediated transfection (Wu and Wu, 1987; 1988). Some of these techniques may be successfully adapted for *in vivo* or *ex vivo* use.

Once the expression polynucleotide has been delivered into the cell, it may be stably integrated into the genome of the recipient cell. This integration may be in the cognate location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle.

One specific embodiment for a method for delivering a protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide of interest into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* as well.

Compositions for use *in vitro* and *in vivo* comprising a "naked" polynucleotide are described in PCT application N° WO 90/11092 (Vical Inc.), and also in PCT application No. WO 95/11307 (Institut Pasteur, INSERM, Université d'Ottawa), as well as in the articles of Tacson *et al.*(1996) and of Huygen *et al.*(1996).

In still another embodiment of the invention, the transfer of a naked polynucleotide of the invention, including a polynucleotide construct of the invention, into cells may be proceeded with a particle bombardment (biolistic), said particles being DNA-coated microprojectiles accelerated to a high velocity allowing them to pierce cell membranes and enter cells without killing them, such as described by Klein *et al.*(1987).

In a further embodiment, the polynucleotide of the invention may be entrapped in a liposome (Ghosh and Bacchawat, 1991; Wong *et al.*, 1980; Nicolau *et al.*, 1987)

In a specific embodiment, the invention provides a composition for the *in vivo* production of the PG-3 protein or polypeptide described herein. It comprises a naked polynucleotide operatively coding for this polypeptide, in solution in a physiologically acceptable carrier, and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

The amount of vector to be injected to the desired host organism varies according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, it may be introduced *in vitro* in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired PG-3 polypeptide or the desired fragment thereof is reintroduced into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

CELL HOSTS

Another object of the invention consists of a host cell that has been transformed or transfected with one of the polynucleotides described herein, and in particular a polynucleotide either comprising a PG-3 regulatory polynucleotide or the coding sequence for the PG-3 polypeptide in a polynucleotide selected from the group consisting of SEQ ID Nos 1 and 2 or a fragment or a variant thereof. Also included are host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above. More particularly, the cell hosts of the present invention can comprise any of the polynucleotides described in the "Genomic Sequences Of The PG3 Gene" section, the "PG-3 cDNA Sequences" section, the "Coding Regions" section, the "Polynucleotide constructs" section, and the "Oligonucleotide Probes And Primers" section.

A further recombinant cell host according to the invention comprises a polynucleotide containing a biallelic marker selected from the group consisting of A1 to A80, and the complements thereof.

An additional recombinant cell host according to the invention comprises any of the vectors described herein, more particularly any of the vectors described in the " Recombinant Vectors" section.

Preferred host cells used as recipients for the expression vectors of the invention are the following:

a) Prokaryotic host cells: *Escherichia coli* strains (*I.E.*DH5-α strain), *Bacillus subtilis*, *Salmonella typhimurium*, and strains from species like *Pseudomonas*, *Streptomyces* and *Staphylococcus*.

b) Eukaryotic host cells: HeLa cells (ATCC N°CCL2; N°CCL2.1; N°CCL2.2), Cv 1 cells (ATCC N°CCL70), COS cells (ATCC N°CRL1650; N°CRL1651), Sf-9 cells (ATCC N°CRL1711), C127 cells (ATCC N° CRL-1804), 3T3 (ATCC N° CRL-6361), CHO (ATCC N° CCL-61), human kidney 293. (ATCC N° 45504; N° CRL-1573) and BHK (ECACC N° 84100501; N° 84111301).

c) Other mammalian host cells.

The PG-3 gene expression in mammalian, and typically human, cells may be rendered defective, or alternatively expression may be provided by the insertion of a PG-3 genomic or cDNA sequence with the replacement of the PG-3 gene counterpart in the genome of an animal cell by a PG-3 polynucleotide according to the invention. These genetic alterations may be generated by homologous recombination events using specific DNA constructs that have been previously described.

One kind of cell hosts that may be used are mammalian zygotes, such as murine zygotes. For example, murine zygotes may undergo microinjection with a purified DNA molecule of interest, for example a purified DNA molecule that has previously been adjusted to a concentration range from 1 ng/ml –for BAC inserts- 3 ng/µl –for P1 bacteriophage inserts- in 10 mM Tris-HCl, pH 7.4, 250 µM EDTA containing 100 mM NaCl, 30 µM spermine, and70 µM spermidine. When the DNA to be microinjected has a large size, polyamines and high salt concentrations can be used in order to avoid mechanical breakage of this DNA, as described by Schedl et al (1993b).

Anyone of the polynucleotides of the invention, including the DNA constructs described herein, may be introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommitted cells of the inner cell mass of pre-implantation blastocysts. Preferred ES cell lines are the following: ES-E14TG2a (ATCC n° CRL-1821), ES-D3 (ATCC n° CRL1934 and n° CRL-11632), YS001 (ATCC n° CRL-11776), 36.5 (ATCC n° CRL-11116). To maintain ES cells in an uncommitted state, they are cultured in the presence of growth inhibited feeder cells which provide the appropriate signals to preserve this embryonic phenotype

and serve as a matrix for ES cell adherence. Preferred feeder cells consist of primary embryonic fibroblasts that are established from tissue of day 13- day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo *et al.*(1993) and are inhibited in growth by irradiation, such as described by Robertson (1987), or by the presence of an inhibitory concentration of LIF, such as described by Pease and Williams (1990).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in the expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

TRANSGENIC ANIMALS

The terms "transgenic animals" or "host animals" are used herein designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention. In one embodiment, the invention encompasses non-human host mammals and animals comprising a recombinant vector of the invention or a PG-3 gene disrupted by homologous recombination with a knock out vector.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a PG-3 coding sequence, a PG-3 regulatory polynucleotide, a polynucleotide construct, or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

Generally, a transgenic animal according the present invention comprises any one of the polynucleotides, the recombinant vectors and the cell hosts described in the present invention. More particularly, the transgenic animals of the present invention can comprise any of the polynucleotides described in the "Genomic Sequences Of The PG3 Gene" section, the "PG-3 cDNA Sequences" section, the "Coding Regions" section, the "Polynucleotide constructs" section, the "Oligonucleotide Probes And Primers" section, the "Recombinant Vectors" section and the "Cell Hosts" section.

A further transgenic animals according to the invention contains in their somatic cells and/or in their germ line cells a polynucleotide comprising a biallelic marker selected from the group consisting of A1 to A80, and the complements thereof.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to cell differentiation, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native PG-3 protein, or alternatively a mutant PG-3 protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of the regulatory polynucleotides of the PG-3 gene, leading to good yields in the synthesis of this protein of interest, and eventually a tissue specific expression of this protein of interest.

The design of the transgenic animals of the invention may be made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to US Patents Nos 4,873,191, issued Oct. 10, 1989; 5,464,764 issued Nov 7, 1995; and 5,789,215, issued Aug 4, 1998; these documents disclosing methods producing transgenic mice.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a PG-3 coding sequence, a PG-3 regulatory polynucleotide or a DNA sequence encoding a PG-3 antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas *et al.*(1987). The cells subjected to electroporation are screened (e.g. by selection via selectable markers, by PCR or by Southern blot analysis) to find positive cells which have integrated the exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour *et al.*(1988).

Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley (1987). The blastocysts are then inserted into a female host animal and allowed to grow to term.

Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood *et al.*(1993) or by Nagy *et al.*(1993), the ES cells being internalized to colonize extensively the blastocyst including the cells which will give rise to the germ line.

The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention.

A further object of the invention consists of recombinant host cells obtained from a transgenic animal described herein. In one embodiment the invention encompasses cells derived from non-human host mammals and animals comprising a recombinant vector of the invention or a PG-3 gene disrupted by homologous recombination with a knock out vector.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989) and Shay *et al.*(1991).

METHODS FOR SCREENING SUBSTANCES INTERACTING WITH A PG-3 POLYPEPTIDE

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to the PG-3 protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for PG-3 or a fragment or variant thereof. These molecules may be used in therapeutic compositions, preferably therapeutic compositions acting against cancer.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of the PG-3 protein is brought into contact with the corresponding purified PG-3 protein, for example the corresponding purified recombinant PG-3 protein produced by a recombinant cell host as described hereinbefore, in order to form a complex between this protein and the putative ligand molecule to be tested.

As an illustrative example, to study the interaction of the PG-3 protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, with drugs or small molecules, such as molecules generated through combinatorial chemistry approaches, the microdialysis coupled to HPLC method described by Wang *et al.* (1997) or the affinity capillary electrophoresis method described by Bush *et al.* (1997).

In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with the PG-3 protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3 may be identified using assays such as the following. The molecule to be tested for binding is labeled with a detectable label, such as a fluorescent , radioactive, or enzymatic tag and placed in contact with immobilized PG-3 protein, or a fragment thereof under conditions which permit specific binding to occur. After removal of non-specifically bound molecules, bound molecules are detected using appropriate means.

Another object of the present invention consists of methods and kits for the screening of candidate substances that interact with PG-3 polypeptide.

The present invention pertains to methods for screening substances of interest that interact with a PG-3 protein or one fragment or variant thereof. By their capacity to bind covalently or non-covalently to a PG-3 protein or to a fragment or variant thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*.

In vitro, said interacting molecules may be used as detection means in order to identify the presence of a PG-3 protein in a sample, preferably a biological sample.

A method for the screening of a candidate substance comprises the following steps :

a) providing a polypeptide consisting of a PG-3 protein or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance;

d) detecting the complexes formed between said polypeptide and said candidate substance.

The invention further concerns a kit for the screening of a candidate substance interacting with the PG-3 polypeptide, wherein said kit comprises:

a) a PG-3 protein having an amino acid sequence selected from the group consisting of the amino acid sequences of SEQ ID No 3 or a peptide fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3;

b) optionally means useful to detect the complex formed between the PG-3 protein or a peptide fragment or a variant thereof and the candidate substance.

In a preferred embodiment of the kit described above, the detection means consist in monoclonal or polyclonal antibodies directed against the PG-3 protein or a peptide fragment or a variant thereof.

Various candidate substances or molecules can be assayed for interaction with a PG-3 polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule consists of a polypeptide, this polypeptide may be the resulting expression product of a phage clone belonging to a phage-based random peptide library, or alternatively the polypeptide may be the resulting expression product of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a PG-3 polypeptide or a fragment or a variant thereof, and optionally means useful to detect the complex formed between the PG-3 polypeptide or

its fragment or variant and the candidate substance. In a preferred embodiment the detection means consist in monoclonal or polyclonal antibodies directed against the corresponding PG-3 polypeptide or a fragment or a variant thereof.

A. Candidate ligands obtained from random peptide libraries

In a particular embodiment of the screening method, the putative ligand is the expression product of a DNA insert contained in a phage vector (Parmley and Smith, 1988). Specifically, random peptide phages libraries are used. The random DNA inserts encode for peptides of 8 to 20 amino acids in length (Oldenburg K.R. *et al.*, 1992; Valadon P., *et al.*, 1996; Lucas A.H., 1994; Westerink M.A.J., 1995; Felici F. *et al.*, 1991). According to this particular embodiment, the recombinant phages expressing a protein that binds to the immobilized PG-3 protein is retained and the complex formed between the PG-3 protein and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the PG-3 protein.

Once the ligand library in recombinant phages has been constructed, the phage population is brought into contact with the immobilized PG-3 protein. Then the preparation of complexes is washed in order to remove the non-specifically bound recombinant phages. The phages that bind specifically to the PG-3 protein are then eluted by a buffer (acid pH) or immunoprecipitated by the monoclonal antibody produced by the hybridoma anti-PG-3, and this phage population is subsequently amplified by an over-infection of bacteria (for example E. coli). The selection step may be repeated several times, preferably 2-4 times, in order to select the more specific recombinant phage clones. The last step consists in characterizing the peptide produced by the selected recombinant phage clones either by expression in infected bacteria and isolation, expressing the phage insert in another host-vector system, or sequencing the insert contained in the selected recombinant phages.

B. Candidate ligands obtained by competition experiments.

Alternatively, peptides, drugs or small molecules which bind to the PG-3 protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, may be identified in competition experiments. In such assays, the PG-3 protein, or a fragment thereof, is immobilized to a surface, such as a plastic plate. Increasing amounts of the peptides, drugs or small molecules are placed in contact with the immobilized PG-3 protein, or a fragment thereof, in the presence of a detectable labeled known PG-3 protein ligand. For example, the PG-3 ligand may be detectably labeled with a fluorescent, radioactive, or enzymatic tag. The ability of the test molecule to bind the PG-3 protein, or a fragment thereof, is determined by measuring the amount of detectably labeled known ligand bound in the presence of the test molecule. A decrease in the amount of known ligand bound to the PG-3 protein, or a fragment thereof, when the test molecule is present indicated that the test molecule is able to bind to the PG-3 protein, or a fragment thereof.

C. Candidate ligands obtained by affinity chromatography.

Proteins or other molecules interacting with the PG-3 protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, can also be found using affinity columns which contain the PG-3 protein, or a fragment thereof. The PG-3 protein, or a fragment thereof, may be attached to the column using conventional techniques including chemical coupling to a suitable column matrix such as agarose, Affi Gel® , or other matrices familiar to those of skill in art. In some embodiments of this method, the affinity column contains chimeric proteins in which the PG-3 protein, or a fragment thereof, is fused to glutathion S transferase (GST). A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Proteins or other molecules interacting with the PG-3 protein, or a fragment thereof, attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen *et al.* (1997). Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

D. Candidate ligands obtained by optical biosensor methods

Proteins interacting with the PG-3 protein, or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3, can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow (1997) and also in Szabo *et al.* (1995). This technique permits the detection of interactions between molecules in real time, without the need of labeled molecules. This technique is based on the surface plasmon resonance (SPR) phenomenon. Briefly, the candidate ligand molecule to be tested is attached to a surface (such as a carboxymethyl dextran matrix). A light beam is directed towards the side of the surface that does not contain the sample to be tested and is reflected by said surface. The SPR phenomenon causes a decrease in the intensity of the reflected light with a specific association of angle and wavelength. The binding of candidate ligand molecules cause a change in the refraction index on the surface, which change is detected as a change in the SPR signal. For screening of candidate ligand molecules or substances that are able to interact with the PG-3 protein, or a fragment thereof, the PG-3 protein, or a fragment thereof, is immobilized onto a surface. This surface consists of one side of a cell through which flows the candidate molecule to be assayed. The binding of the candidate molecule on the PG-3 protein, or a fragment thereof, is detected as a change of the SPR signal. The candidate molecules tested may be proteins, peptides, carbohydrates, lipids, or small molecules generated by combinatorial chemistry. This technique may also be performed by immobilizing eukaryotic or prokaryotic cells or lipid vesicles exhibiting an endogenous or a recombinantly expressed PG-3 protein at their surface.

The main advantage of the method is that it allows the determination of the association rate between the PG-3 protein and molecules interacting with the PG-3 protein. It is thus possible to

select specifically ligand molecules interacting with the PG-3 protein, or a fragment thereof, through strong or conversely weak association constants.

E. Candidate ligands obtained through a two-hybrid screening assay.

The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields and Song, 1989), and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in the US Patent N° US 5,667,973 and the US Patent N° 5,283,173.

The general procedure of library screening by the two-hybrid assay may be performed as described by Harper *et al.* (1993) or as described by Cho *et al.* (1998) or also Fromont-Racine *et al.* (1997).

The bait protein or polypeptide consists of a PG-3 polypeptide or a fragment comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3.

More precisely, the nucleotide sequence encoding the PG-3 polypeptide or a fragment or variant thereof is fused to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example pAS2 or pM3.

Then, a human cDNA library is constructed in a specially designed vector, such that the human cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcriptional domain of the GAL4 protein. Preferably, the vector used is the pACT vector. The polypeptides encoded by the nucleotide inserts of the human cDNA library are termed "pray" polypeptides.

A third vector contains a detectable marker gene, such as beta galactosidase gene or CAT gene that is placed under the control of a regulation sequence that is responsive to the binding of a complete Gal4 protein containing both the transcriptional activation domain and the DNA binding domain. For example, the vector pG5EC may be used.

Two different yeast strains are also used. As an illustrative but non limiting example the two different yeast strains may be the followings :

- Y190, the phenotype of which is (*MATa, Leu2-3, 112 ura3-12, trp1-901, his3-D200, ade2-101, gal4Dgal180D URA3 GAL-LacZ, LYS GAL-HIS3, cyh*^{r});
- Y187, the phenotype of which is (*MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112 URA3 GAL-lacZmet*$^-$), which is the opposite mating type of Y190.

Briefly, 20 µg of pAS2/PG-3 and 20 µg of pACT-cDNA library are co-transformed into yeast strain Y190. The transformants are selected for growth on minimal media lacking histidine, leucine and tryptophan, but containing the histidine synthesis inhibitor 3-AT (50 mM). Positive colonies are screened for beta galactosidase by filter lift assay. The double positive colonies (*His*$^+$, *beta-gal*$^+$) are then grown on plates lacking histidine, leucine, but containing tryptophan and cycloheximide (10 mg/ml) to select for loss of pAS2/PG-3 plasmids bu retention of pACT-cDNA

library plasmids. The resulting Y190 strains are mated with Y187 strains expressing PG-3 or non-related control proteins; such as cyclophilin B, lamin, or SNF1, as *Gal4* fusions as described by Harper *et al.* (1993) and by Bram *et al.* (1993), and screened for beta galactosidase by filter lift assay. Yeast clones that are *beta gal-* after mating with the control *Gal4* fusions are considered false positives.

In another embodiment of the two-hybrid method according to the invention, interaction between the PG-3 or a fragment or variant thereof with cellular proteins may be assessed using the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech), nucleic acids encoding the PG-3 protein or a portion thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. A desired cDNA, preferably human cDNA, is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain interaction between PG-3 and the protein or peptide encoded by the initially selected cDNA insert.

METHOD FOR SCREENING SUBSTANCES INTERACTING WITH THE REGULATORY SEQUENCES OF THE PG-3 GENE.

The present invention also concerns a method for screening substances or molecules that are able to interact with the regulatory sequences of the PG-3 gene, such as for example promoter or enhancer sequences.

Nucleic acids encoding proteins which are able to interact with the regulatory sequences of the PG-3 gene, more particularly a nucleotide sequence selected from the group consisting of the polynucleotides of the 5' and 3' regulatory region or a fragment or variant thereof, and preferably a variant comprising one of the biallelic markers of the invention, may be identified by using a one-hybrid system, such as that described in the booklet enclosed in the Matchmaker One-Hybrid System kit from Clontech (Catalog Ref. n° K1603-1). Briefly, the target nucleotide sequence is cloned upstream of a selectable reporter sequence and the resulting DNA construct is integrated in the yeast genome (*Saccharomyces cerevisiae*). The yeast cells containing the reporter sequence in their genome are then transformed with a library consisting of fusion molecules between cDNAs encoding candidate proteins for binding onto the regulatory sequences of the PG-3 gene and sequences encoding the activator domain of a yeast transcription factor such as GAL4. The recombinant yeast cells are plated in a culture broth for selecting cells expressing the reporter sequence. The recombinant yeast cells thus selected contain a fusion protein that is able to bind

onto the target regulatory sequence of the PG-3 gene. Then, the cDNAs encoding the fusion proteins are sequenced and may be cloned into expression or transcription vectors *in vitro*. The binding of the encoded polypeptides to the target regulatory sequences of the PG-3 gene may be confirmed by techniques familiar to the one skilled in the art, such as gel retardation assays or DNAse protection assays.

Gel retardation assays may also be performed independently in order to screen candidate molecules that are able to interact with the regulatory sequences of the PG-3 gene, such as described by Fried and Crothers (1981), Garner and Revzin (1981) and Dent and Latchman (1993). These techniques are based on the principle according to which a DNA fragment which is bound to a protein migrates slower than the same unbound DNA fragment. Briefly, the target nucleotide sequence is labeled. Then the labeled target nucleotide sequence is brought into contact with either a total nuclear extract from cells containing transcription factors, or with different candidate molecules to be tested. The interaction between the target regulatory sequence of the PG-3 gene and the candidate molecule or the transcription factor is detected after gel or capillary electrophoresis through a retardation in the migration.

METHOD FOR SCREENING LIGANDS THAT MODULATE THE EXPRESSION OF THE PG-3 GENE.

Another subject of the present invention is a method for screening molecules that modulate the expression of the PG-3 protein. Such a screening method comprises the steps of:

a) cultivating a prokaryotic or an eukaryotic cell that has been transfected with a nucleotide sequence encoding the PG-3 protein or a variant or a fragment thereof, placed under the control of its own promoter;

b) bringing into contact the cultivated cell with a molecule to be tested;

c) quantifying the expression of the PG-3 protein or a variant or a fragment thereof.

In an embodiment, the nucleotide sequence encoding the PG-3 protein or a variant or a fragment thereof comprises an allele of at least one of the biallelic markers A1 to A80, and the complements thereof.

Using DNA recombination techniques well known by the one skill in the art, the PG-3 protein encoding DNA sequence is inserted into an expression vector, downstream from its promoter sequence. As an illustrative example, the promoter sequence of the PG-3 gene is contained in the nucleic acid of the 5' regulatory region.

The quantification of the expression of the PG-3 protein may be realized either at the mRNA level or at the protein level. In the latter case, polyclonal or monoclonal antibodies may be used to quantify the amounts of the PG-3 protein that have been produced, for example in an ELISA or a RIA assay.

In a preferred embodiment, the quantification of the PG-3 mRNA is realized by a quantitative PCR amplification of the cDNA obtained by a reverse transcription of the total mRNA of the cultivated PG-3 -transfected host cell, using a pair of primers specific for PG-3.

The present invention also concerns a method for screening substances or molecules that are able to increase, or in contrast to decrease, the level of expression of the PG-3 gene. Such a method may allow the one skilled in the art to select substances exerting a regulating effect on the expression level of the PG-3 gene and which may be useful as active ingredients included in pharmaceutical compositions for treating patients suffering from cancer.

Thus, another aspect of the present invention is a method for screening a candidate substance or molecule for the ability to modulate the expression of the PG-3 gene, comprising the following steps:

a) providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid comprises a nucleotide sequence of the 5' regulatory region or a biologically active fragment or variant thereof located upstream of a polynucleotide encoding a detectable protein;

b) obtaining a candidate substance; and

c) determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

In a further embodiment, the nucleic acid comprising the nucleotide sequence of the 5' regulatory region or a biologically active fragment or variant thereof also includes a 5'UTR region of the PG-3 cDNA of SEQ ID No 2, or one of its biologically active fragments or variants thereof.

Among the preferred polynucleotides encoding a detectable protein, there may be cited polynucleotides encoding beta galactosidase, green fluorescent protein (GFP) and chloramphenicol acetyl transferase (CAT).

The invention also pertains to kits useful for performing the herein described screening method. Preferably, such kits comprise a recombinant vector that allows the expression of a nucleotide sequence of the 5' regulatory region or a biologically active fragment or variant thereof located upstream and operably linked to a polynucleotide encoding a detectable protein or the PG-3 protein or a fragment or a variant thereof.

In another embodiment of a method for the screening of a candidate substance or molecule for the ability to modulate the expression of the PG-3 gene, the method comprises the following steps:

a) providing a recombinant host cell containing a nucleic acid, wherein said nucleic acid comprises a 5'UTR sequence of the PG-3 cDNA of SEQ ID No 2, or one of its biologically active fragments or variants, the 5'UTR sequence or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein;

b) obtaining a candidate substance; and

c) determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

In a specific embodiment of the above screening method, the nucleic acid that comprises a nucleotide sequence selected from the group consisting of the 5'UTR sequence of the PG-3 cDNA of SEQ ID No 2 or one of its biologically active fragments or variants, includes a promoter sequence which is endogenous with respect to the PG-3 5'UTR sequence.

In another specific embodiment of the above screening method, the nucleic acid that comprises a nucleotide sequence selected from the group consisting of the 5'UTR sequence of the PG-3 cDNA of SEQ ID No 2 or one of its biologically active fragments or variants, includes a promoter sequence which is exogenous with respect to the PG-3 5'UTR sequence defined therein.

In a further preferred embodiment, the nucleic acid comprising the 5'-UTR sequence of the PG-3 cDNA or SEQ ID No 2 or the biologically active fragments thereof includes a biallelic marker selected from the group consisting of A1 to A80 or the complements thereof.

The invention further encompasses a kit for the screening of a candidate substance for the ability to modulate the expression of the PG-3 gene, wherein said kit comprises a recombinant vector that comprises a nucleic acid including a 5'UTR sequence of the PG-3 cDNA of SEQ ID No 2, or one of their biologically active fragments or variants, the 5'UTR sequence or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein.

For the design of suitable recombinant vectors useful for performing the screening methods described above, the section of the present specification wherein the preferred recombinant vectors of the invention are detailed is pertinent.

Expression levels and patterns of PG-3 may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277. Briefly, the PG-3 cDNA or the PG-3 genomic DNA described above, or fragments thereof, is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the PG-3 insert comprises at least 100 or more consecutive nucleotides of the genomic DNA sequence or the cDNA sequences. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (*i.e.* biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridization is performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0. 4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (*i.e.* RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

Quantitative analysis of PG-3 gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of a plurality of nucleic acids of sufficient length to permit specific detection of expression of mRNAs capable of hybridizing thereto. For example, the arrays may contain a plurality of nucleic acids derived from genes whose expression levels are to be assessed. The arrays may include the PG-3 genomic DNA, the PG-3 cDNA sequences or the sequences complementary thereto or fragments thereof, particularly those comprising at least one of the biallelic markers according the present invention, preferably at least one of the biallelic markers A1 to A80. Preferably, the fragments are at least 15 nucleotides in length. In other embodiments, the fragments are at least 25 nucleotides in length. In some embodiments, the fragments are at least 50 nucleotides in length. More preferably, the fragments are at least 100 nucleotides in length. In another preferred embodiment, the fragments are more than 100 nucleotides in length. In some embodiments the fragments may be more than 500 nucleotides in length.

For example, quantitative analysis of PG-3 gene expression may be performed with a complementary DNA microarray as described by Schena *et al.*(1995 and 1996). Full length PG-3 cDNAs or fragments thereof are amplified by PCR and arrayed from a 96-well microtiter plate onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0. 2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0. 2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm^2 microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1X SSC/0. 2% SDS), then for 10 min at room temperature in high stringency wash buffer (0. 1X SSC/0. 2% SDS). Arrays are scanned in 0. 1X SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of PG-3 gene expression may also be performed with full length PG-3 cDNAs or fragments thereof in complementary DNA arrays as described by Pietu *et al.*(1996). The full length PG-3 cDNA or fragments thereof is PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis using the PG-3 genomic DNA, the PG-3 cDNA, or fragments thereof can be done through high density nucleotide arrays as described by Lockhart *et*

al.(1996) and Sosnowski *et al.*(1997). Oligonucleotides of 15-50 nucleotides from the sequences of the PG-3 genomic DNA, the PG-3 cDNA sequences particularly those comprising at least one of biallelic markers according the present invention, preferably at least one biallelic marker selected from the group consisting of A1 to A80, or the sequences complementary thereto, are synthesized directly on the chip (Lockhart *et al.*, supra) or synthesized and then addressed to the chip (Sosnowski *et al.*, supra). Preferably, the oligonucleotides are about 20 nucleotides in length.

PG-3 cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart *et al.*, supra and application of different electric fields (Sosnowski *et al.*, 1997), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of PG-3 mRNA.

METHODS FOR INHIBITING THE EXPRESSION OF A PG-3 GENE

Other therapeutic compositions according to the present invention comprise advantageously an oligonucleotide fragment of the nucleic sequence of PG-3 as an antisense tool or a triple helix tool that inhibits the expression of the corresponding PG-3 gene. A preferred fragment of the nucleic sequence of PG-3 comprises an allele of at least one of the biallelic markers A1 to A80.

Antisense Approach

Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel *et al.*(1995).

Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of the PG-3 mRNA. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of PG-3 that contains either the translation initiation codon ATG or a splicing donor or acceptor site.

The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the PG-3 mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et al.*, (1986) and Izant and Weintraub, (1984).

In some strategies, antisense molecules are obtained by reversing the orientation of the PG-3 coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript.

Another approach involves transcription of PG-3 antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, suitable antisense strategies are those described by Rossi *et al.*(1991), in the International Applications Nos. WO 94/23026, WO 95/04141, WO 92/18522 and in the European Patent Application No. EP 0 572 287 A2.

An alternative to the antisense technology that is used according to the present invention consists in using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme consists of (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present invention are prepared as described by Sczakiel *et al.*(1995).

Triple Helix Approach

The PG-3 genomic DNA may also be used to inhibit the expression of the PG-3 gene based on intracellular triple helix formation.

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene.

Similarly, a portion of the PG-3 genomic DNA can be used to study the effect of inhibiting PG-3 transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the PG-3 genomic DNA are contemplated within the scope of this invention.

To carry out gene therapy strategies using the triple helix approach, the sequences of the PG-3 genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting PG-3 expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting PG-3 expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the PG-3 gene.

The oligonucleotides can be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced PG-3 expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the PG-3 gene in cells which have been treated with the oligonucleotide.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced in vivo using the techniques described above in the antisense approach at a dosage calculated based on the in vitro results, as described in antisense approach.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin *et al.*(1989), which is hereby incorporated by this reference.

COMPUTER-RELATED EMBODIMENTS

As used herein the term "nucleic acid codes of the invention" encompass the nucleotide sequences comprising, consisting essentially of, or consisting of any one of the following: a) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825; b) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof; and, c) a nucleotide sequence complementary to any one of the preceding nucleotide sequences.

The "nucleic acid codes of the invention" further encompass nucleotide sequences homologous to: a) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825; b) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 2 or the complements thereof; and, c) sequences complementary to all of the preceding sequences. Homologous sequences refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% homology to these contiguous spans. Homology may be determined using any method described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also may include RNA sequences in which uridines replace the thymines in the nucleic acid codes of the invention. It will be appreciated that the nucleic acid codes of the invention can be represented in the traditional single character format (See the inside back cover of Stryer,

Lubert. 1995) or in any other format or code which records the identity of the nucleotides in a sequence.

As used herein the term "polypeptide codes of the invention" encompass the polypeptide sequences comprising a contiguous span of at least 6, 8, 10, 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID No 3. It will be appreciated that the polypeptide codes of the invention can be represented in the traditional single character format or three letter format (See the inside back cover of Stryer, Lubert.) or in any other format or code which records the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention and polypeptide codes of the invention can be stored, recorded, and manipulated on any medium which can be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention, or one or more of the polypeptide codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 nucleic acid codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of the invention.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 1. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the computer system 100 is a Sun Enterprise 1000 server (Sun Microsystems, Palo Alto, CA). The computer system 100 preferably includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq or International Business Machines.

Preferably, the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the computer system 100 further includes one or more data retrieving device 118 for reading the data stored on the internal data storage devices 110.

The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, etc. In some embodiments, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic and/or the data from the data storage component once inserted in the data retrieving device.

The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution.

In some embodiments, the computer system 100 may further comprise a sequence comparer for comparing the above-described nucleic acid codes of the invention or the polypeptide codes of the invention stored on a computer readable medium to reference nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system 100 to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

Figure 2 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK, PIR OR SWISSPROT that is available through the Internet.

99-12767	A38	99-12767-189	Intron J-K	C	T	123277		
99-12767	A39	99-12767-380	Intron J-K	A	G	123468		
4-80	A40	4-80-328	Intron J-K	C	T	126738		
4-36	A41	4-36-384	Intron J-K	G	C	128210		
4-36	A42	4-36-264	Intron J-K	A	G	128330		
4-36	A43	4-36-261	Intron J-K	A	C	128333		
4-35	A44	4-35-333	Intron J-K	A	C	128594		
4-35	A45	4-35-240	Intron J-K	G	C	128687		
4-35	A46	4-35-173	Intron J-K	A	T	128754		
4-35	A47	4-35-133	Intron J-K	C	T	128794		
99-12771	A48	99-12771-59	Intron J-K	G	T	130805		
99-12774	A49	99-12774-334	Intron J-K	A	C	133206		
99-12776	A50	99-12776-358	Intron J-K	A	G	135386		
99-12781	A51	99-12781-113	Intron J-K	A	G	139389		
4-104	A52	4-104-298	Intron J-K	G	C	157535		
4-104	A53	4-104-254	Intron J-K	A	G	157579		
4-104	A54	4-104-250	Intron J-K	C	T	157583		
4-104	A55	4-104-214	Intron J-K	A	G	157619		
99-12818	A56	99-12818-289	Intron J-K	C	T	172980		
99-24807	A57	99-24807-271	Intron J-K	C	T	180622		
99-24807	A58	99-24807-84	Intron J-K	A	G	180809		
99-12831	A59	99-12831-157	Intron J-K	A	G	190334		
99-12831	A60	99-12831-241	Intron J-K	C	T	190418		
99-12832	A61	99-12832-387	Intron J-K	C	T	191397		
99-12836	A62	99-12836-30	Intron J-K	G	C	195128		
99-12844	A63	99-12844-262	Intron J-K	G	C	203846		
4-24	A64	4-24-74	Intron J-K	C	T	210151		
4-24	A65	4-24-246	Intron J-K	C	T	210321		
4-24	A66	4-24-314	Intron J-K	G	C	210389		
4-27	A67	4-27-190	Intron J-K	A	G	211168		
5-400	A68	5-400-145	Intron J-K	A	G	215996		
5-400	A69	5-400-149	Intron J-K	G	C	216000		
5-400	A70	5-400-175	Exon K	C	T	216026	2283	742 = S
5-400	A71	5-400-231	Exon K	C	T	216082	2339	761 = A or V
5-400	A72	5-400-367	Exon K	A	C	216218	2475	806 = A
99-12852	A73	99-12852-110	Intron K-L	G	T	216322		
99-12852	A74	99-12852-325	Intron K-L	A	G	216537		
4-37	A75	4-37-326	Intron K-L	A	C	221649		
4-37	A76	4-37-107	Intron K-L	A	G	221867		
5-270	A77	5-270-92	Intron K-L	G	C	225645		
99-12860	A78	99-12860-47	Intron K-L	A	G	229387		
99-12860	A79	99-12860-57	Intron K-L	A	T	229397		
5-402	A80	5-402-144	Ex n L	C	T	237555	2539	828 = P or S

BM refers to "biallelic marker". All1 and all2 refer respectively to allele 1 and allele 2 of th biallelic marker.

The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters that control whether gaps or other features are introduced into a sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately to the decision state 218 in order to determine if any other sequences were available in the database for comparison.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a polypeptide code of the invention, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of the invention and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the nucleic acid code of the invention and polypeptide codes of the invention or it may identify structural motifs in sequences which are compared to these

nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or polypeptide codes of the invention.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes of the invention through the use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences.

Figure 3 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it should be in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there aren't any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of the invention differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of the invention. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain one or more single nucleotide polymorphisms (SNP) with respect to a reference nucleotide sequence. These single nucleotide polymorphisms may each comprise a single base substitution, insertion, or deletion.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of the invention and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of the invention and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms The method may be implemented by the computer systems described above and the method illustrated in Figure 3. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention.

An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention. In one embodiment, the identifier may comprise a program which identifies an open reading frame in the cDNAs codes of the invention.

Figure 4 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a

memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group (www.gcg.com).

Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether move features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence.

It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

In another embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of the invention. In some embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg *et al.*, U.S. Patent No. 5,436,850 issued July 25, 1995). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of the invention. (See e.g., Srinivasan, *et al.*, U.S. Patent No. 5,557,535 issued September 17, 1996). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini *et al.*, (1997)). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA.

According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi *et al.*, (1997)).

The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of the invention.

Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of the invention or the polypeptide codes of the invention comprising reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or the polypeptide codes of the invention through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

The nucleic acid codes of the invention or the polypeptide codes of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, they may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of the invention. The following list is intended not to limit the invention but to provide guidance to programs and databases which are

useful with the nucleic acid codes of the invention or the polypeptide codes of the invention. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, 1990), FASTA (Pearson and Lipman, 1988), FASTDB (Brutlag *et al.*, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius^2.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, the Genseqn database and the Genseqp databases. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

Throughout this application, various publications, patents and published patent applications are cited. The disclosures of these publications, patents and published patent specification referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the sate of the art to which this invention pertains.

EXAMPLES

EXAMPLE 1

IDENTIFICATION OF BIALLELIC MARKERS - DNA EXTRACTION

Donors were unrelated and healthy. They presented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the detection of the biallelic markers.

30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume: 10 mM Tris pH7.6; 5 mM $MgCl_2$; 10 mM NaCl). The

solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0 4 M
- 200 µl SDS 10%
- 500 µl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA).

To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below.

The pool was constituted by mixing equivalent quantities of DNA from each individual.

EXAMPLE 2

IDENTIFICATION OF BIALLELIC MARKERS: AMPLIFICATION OF GENOMIC DNA BY PCR

The amplification of specific genomic sequences of the DNA samples of example 1 was carried out on the pool of DNA obtained previously. In addition, 50 individual samples were similarly amplified.

PCR assays were performed using the following protocol:

Final volume	25 µl
DNA	2 ng/µl
$MgCl_2$	2 mM
dNTP (each)	200 µM
primer (each)	2.9 ng/µl
Ampli Taq Gold DNA polymerase	0.05 unit/µl
PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl)	1x

Each pair of first primers was designed using the sequence information of the PG-3 gene disclosed herein and the OSP software (Hillier & Green, 1991). This first pair of primers was about

20 nucleotides in length and had the sequences discl sed in Table 1 in the columns labeled PU and RP.

Table 1

Amplicon	Position range of the amplicon in SEQ ID No:1		PU primer name	Position range of amplification primer in SEQ ID No:1		RP primer name	Complementary position range of amplification primer in SEQ ID No:1	
5-390	1823	2125	B1	1823	1840	C1	2108	2125
5-391	4559	4908	B2	4559	4577	C2	4891	4908
5-392	10007	10430	B3	10007	10025	C3	10411	10430
4-59	39556	39970	B4	39556	39574	C4	39953	39970
4-58	39877	40259	B5	39877	39896	C5	40242	40259
4-54	41137	41581	B6	41137	41154	C6	41564	41581
4-51	42122	42543	B7	42122	42141	C7	42526	42543
99-86	67289	67741	B8	67289	67309	C8	67724	67741
4-88	69182	69626	B9	69182	69200	C9	69609	69626
5-397	72698	73117	B10	72698	72715	C10	73099	73117
5-398	75858	76306	B11	75858	75877	C11	76289	76306
99-12738	81006	81485	B12	81006	81025	C12	81466	81485
99-109	83564	84007	B13	83564	83582	C13	83990	84007
99-12749	91743	92142	B14	91743	91763	C14	92123	92142
4-21	95196	95619	B15	95196	95214	C15	95600	95619
4-23	95865	96229	B16	95865	95882	C16	96210	96229
99-12753	97261	97747	B17	97261	97278	C17	97728	97747
5-364	97831	98275	B18	97831	97849	C18	98256	98275
99-12755	98638	99131	B19	98638	98656	C19	99111	99131
4-87	103376	103818	B20	103376	103395	C20	103801	103818
99-12757	104081	104636	B21	104081	104100	C21	104619	104636
99-12758	106272	106799	B22	106272	106291	C22	106780	106799
4-105	108200	108412	B23	108200	108218	C23	108390	108412
4-45	108223	108520	B24	108223	108246	C24	108499	108520
4-44	109123	109471	B25	109123	109142	C25	109454	109471
4-86	114217	114663	B26	114217	114234	C26	114646	114663
4-84	115630	116049	B27	115630	115647	C27	116031	116049
99-78	121991	122401	B28	121991	122011	C28	122384	122401
99-12767	123089	123583	B29	123089	123106	C29	123565	123583
4-80	126711	127065	B30	126711	126729	C30	127048	127065
4-36	128162	128590	B31	128162	128179	C31	128573	128590
4-35	128480	128926	B32	128480	128497	C32	128909	128926
99-12771	130747	131273	B33	130747	130764	C33	131254	131273
99-12774	132873	133325	B34	132873	132892	C34	133305	133325
99-12776	135029	135478	B35	135029	135048	C35	135458	135478
99-12781	139277	139742	B36	139277	139296	C36	139724	139742
4-104	157181	157832	B37	157181	157199	C37	157814	157832
99-12818	172692	173091	B38	172692	172709	C38	173072	173091

99-24807	180248	180892	**B39**	180248	180268	**C39**	180874	180892
99-12827	184662	185156	**B40**	184662	184680	**C40**	185138	185156
99-12831	190178	190663	**B41**	190178	190196	**C41**	190643	190663
99-12832	191011	191460	**B42**	191011	191030	**C42**	191441	191460
99-12836	195099	195587	**B43**	195099	195116	**C43**	195568	195587
99-12844	203585	204115	**B44**	203585	203602	**C44**	204095	204115
4-24	210079	210495	**B45**	210079	210096	**C45**	210476	210495
4-27	210979	211401	**B46**	210979	210996	**C46**	211382	211401
5-400	215852	216271	**B47**	215852	215870	**C47**	216253	216271
99-12852	216213	216728	**B48**	216213	216231	**C48**	216708	216728
4-37	221530	221973	**B49**	221530	221549	**C49**	221956	221973
5-270	225554	225845	**B50**	225554	225572	**C50**	225827	225845
99-12860	229341	229790	**B51**	229341	229359	**C51**	229770	229790
5-402	237412	237766	**B52**	237412	237429	**C52**	237747	237766

Preferably, the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification which was useful for sequencing.

Primers PU contain the following additional PU 5' sequence: TGTAAAACGACGGCCAGT; primers RP contain the following RP 5' sequence: CAGGAAACAGCTATGACC. The primer containing the additional PU 5' sequence is listed in SEQ ID No 4. The primer containing the additional RP 5' sequence is listed in SEQ ID No 5.

The synthesis of these primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

DNA amplification was performed on a Genius II thermocycler. After heating at 95°C for 10 min, 40 cycles were performed. Each cycle comprised: 30 sec at 95°C, 54°C for 1 min, and 30 sec at 72°C. For final elongation, 10 min at 72°C ended the amplification. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

EXAMPLE 3

IDENTIFICATION OF BIALLELIC MARKERS - SEQUENCING OF AMPLIFIED GENOMIC DNA AND IDENTIFICATION OF POLYMORPHISMS

The sequencing of the amplified DNA obtained in example 2 was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software (2.1.2 version)).

The sequence data were further evaluated to detect the presence of biallelic markers within the amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position as described previously.

In the 52 fragments of amplification, 80 biallelic markers were detected. The localization of these biallelic markers are as shown in Table 2.

Table 2

Amplicon	BM	Marker name	Localization in PG-3 gene	Polymorphism		BM position in SEQ ID		Position of amino acid in SEQ ID No:3
				all1	all2	No:1	No:2	
5-390	A1	5-390-177	5' regulatory	G	C	1999		
5-391	A2	5-391-43	Intron A-B	A	G	4601		
5-392	A3	5-392-222	Exon C	G	T	10228	285	76 = V
5-392	A4	5-392-280	Intron C-D	G	T	10286		
5-392	A5	5-392-364	Intron C-D	G	-	10370		
4-59	A6	4-58-318	Exon T	G	T	39944	968	304 = R or I
4-58	A7	4-58-289	Exon T	G	C	39973	997	314 = H or D
4-54	A8	4-54-199	Intron T-G	A	C	41385		
4-54	A9	4-54-180	Intron T-G	A	C	41404		
4-51	A10	4-51-312	Intron T-G	G	C	42232		
99-86	A11	99-86-266	Intron G-H	A	G	67475		
4-88	A12	4-88-107	Intron G-H	A	G	69521		
5-397	A13	5-397-141	Intron G-H	G	T	72838		
5-398	A14	5-398-203	Exon I .	A	C	76060	2102	682 = T or N
99-12738	A15	99-12738-248	Intron I-J	A	C	81253		
99-109	A16	99-109-358	Intron I-J	A	C	83921		
99-12749	A17	99-12749-175	Intron I-J	C	T	91917		
4-21	A18	4-21-154	Intron J-K	C	T	95349		
4-21	A19	4-21-317	Intron J-K	G	T	95511		
4-23	A20	4-23-326 .	Intron J-K	A	G	96190		
99-12753	A21	99-12753-34	Intron J-K	A	T	97294		
5-364	A22	5-364-252	Intron J-K	G	T	98024		
99-12755	A23	99-12755-280	Intron J-K	A	G	98914		
99-12755	A24	99-12755-329	Intron J-K	A	C	98963		
4-87	A25	4-87-212	Intron J-K	A	G	103593		
99-12757	A26	99-12757-318	Intron J-K	C	T	104398		
99-12758	A27	99-12758-102	Intron J-K	A	G	106373		
99-12758	A28	99-12758-136	Intron J-K	C	T	106407		
4-105	A29	4-105-98	Intron J-K	A	G	108315		
4-105	A30	4-105-86	Intron J-K	A	G	108327		
4-45	A31	4-45-49	Intron J-K	C	T	108472		
4-44	A32	4-44-277	Intron J-K	C	T	109196		
4-86	A33	4-86-60	Intron J-K	G	C	114604		
4-84	A34	4-84-334	Intron J-K	A	G	115716		
99-78	A35	99-78-321	Intron J-K	A	T	122083		
99-12767	A36	99-12767-36	Intr n J-K	G	C	123124		
99-12767	A37	99-12767-143	Intron J-K	C	T	123231		

Table 3

BM	Marker name	P siti n range of probes in SEQ ID No 1		Probes
A1	5-390-177	1987	2011	P1
A2	5-391-43	4589	4613	P2
A3	5-392-222	10216	10240	P3
A4	5-392-280	10274	10298	P4
A6	4-58-318	39932	39956	P6
A7	4-58-289	39961	39985	P7
A8	4-54-199	41373	41397	P8
A9	4-54-180	41392	41416	P9
A10	4-51-312	42220	42244	P10
A11	99-86-266	67463	67487	P11
A12	4-88-107	69509	69533	P12
A13	5-397-141	72826	72850	P13
A14	5-398-203	76048	76072	P14
A15	99-12738-248	81241	81265	P15
A16	99-109-358	83909	83933	P16
A17	99-12749-175	91905	91929	P17
A18	4-21-154	95337	95361	P18
A19	4-21-317	95499	95523	P19
A20	4-23-326	96178	96202	P20
A21	99-12753-34	97282	97306	P21
A22	5-364-252	98012	98036	P22
A23	99-12755-280	98902	98926	P23
A24	99-12755-329	98951	98975	P24
A25	4-87-212	103581	103605	P25
A26	99-12757-318	104386	104410	P26
A27	99-12758-102	106361	106385	P27
A28	99-12758-136	106395	106419	P28
A29	4-105-98	108303	108327	P29
A30	4-105-86	108315	108339	P30
A31	4-45-49	108460	108484	P31
A32	4-44-277	109184	109208	P32
A33	4-86-60	114592	114616	P33
A34	4-84-334	115704	115728	P34
A35	99-78-321	122071	122095	P35
A36	99-12767-36	123112	123136	P36
A37	99-12767-143	123219	123243	P37
A38	99-12767-189	123265	123289	P38
A39	99-12767-380	123456	123480	P39
A40	4-80-328	126726	126750	P40
A41	4-36-384	128198	128222	P41
A42	4-36-264	128318	128342	P42
A43	4-36-261	128321	128345	P43
A44	4-35-333	128582	128606	P44

A45	4-35-240	128675	128699	P45
A46	4-35-173	128742	128766	P46
A47	4-35-133	128782	128806	P47
A48	99-12771-59	130793	130817	P48
A49	99-12774-334	133194	133218	P49
A50	99-12776-358	135374	135398	P50
A51	99-12781-113	139377	139401	P51
A52	4-104-298	157523	157547	P52
A53	4-104-254	157567	157591	P53
A54	4-104-250	157571	157595	P54
A55	4-104-214	157607	157631	P55
A56	99-12818-289	172968	172992	P56
A57	99-24807-271	180610	180634	P57
A58	99-24807-84	180797	180821	P58
A59	99-12831-157	190322	190346	P59
A60	99-12831-241	190406	190430	P60
A61	99-12832-387	191385	191409	P61
A62	99-12836-30	195116	195140	P62
A63	99-12844-262	203834	203858	P63
A64	4-24-74	210139	210163	P64
A65	4-24-246	210309	210333	P65
A66	4-24-314	210377	210401	P66
A67	4-27-190	211156	211180	P67
A68	5-400-145	215984	216008	P68
A69	5-400-149	215988	216012	P69
A70	5-400-175	216014	216038	P70
A71	5-400-231	216070	216094	P71
A72	5-400-367	216206	216230	P72
A73	99-12852-110	216310	216334	P73
A74	99-12852-325	216525	216549	P74
A75	4-37-326	221637	221661	P75
A76	4-37-107	221855	221879	P76
A77	5-270-92	225633	225657	P77
A78	99-12860-47	229375	229399	P78
A79	99-12860-57	229385	229409	P79
A80	5-402-144	237543	237567	P80

EXAMPLE 4

VALIDATION OF THE POLYMORPHISMS THROUGH MICROSEQUENCING

The biallelic markers identified in example 3 were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out for each individual DNA sample described in Example 1.

Amplification from genomic DNA of individuals was performed by PCR as described above for the detection of the biallelic markers with the same set of PCR primers (Table 1).

The preferred primers used in microsequencing were about 19 nucleotides in length and hybridized just upstream of the considered polymorphic base. According to the invention, the primers used in microsequencing are detailed in Table 4.

Table 4

Marker name	BM	Mis 1	Position range of microsequencing primer mis 1 in SEQ ID No 1		Mis 2	Complementary position range of microsequencing primer mis. 2 in SEQ ID No 1	
5-390-177	A1	D1	1980	1998	E1	2000	2018
5-391-43	A2	D2	4582	4600	E2	4602	4620
5-392-222	A3	D3	10209	10227	E3	10229	10247
5-392-280	A4	D4	10267	10285	E4	10287	10305
4-58-318	A6	D6	39925	39943	E6	39945	39963
4-58-289	A7	D7	39954	39972	E7	39974	39992
4-54-199	A8	D8	41366	41384	E8	41386	41404
4-54-180	A9	D9	41385	41403	E9	41405	41423
4-51-312	A10	D10	42213	42231	E10	42233	42251
99-86-266	A11	D11	67456	67474	E11	67476	67494
4-88-107	A12	D12	69502	69520	E12	69522	69540
5-397-141	A13	D13	72819	72837	E13	72839	72857
5-398-203	A14	D14	76041	76059	E14	76061	76079
99-12738-248	A15	D15	81234	81252	E15	81254	81272
99-109-358	A16	D16	83902	83920	E16	83922	83940
99-12749-175	A17	D17	91898	91916	E17	91918	91936
4-21-154	A18	D18	95330	95348	E18	95350	95368
4-21-317	A19	D19	95492	95510	E19	95512	95530
4-23-326	A20	D20	96171	96189	E20	96191	96209
99-12753-34	A21	D21	97275	97293	E21	97295	97313
5-364-252	A22	D22	98005	98023	E22	98025	98043
99-12755-280	A23	D23	98895	98913	E23	98915	98933
99-12755-329	A24	D24	98944	98962	E24	98964	98982
4-87-212	A25	D25	103574	103592	E25	103594	103612
99-12757-318	A26	D26	104379	104397	E26	104399	104417
99-12758-102	A27	D27	106354	106372	E27	106374	106392
99-12758-136	A28	D28	106388	106406	E28	106408	106426
4-105-98	A29	D29	108296	108314	E29	108316	108334

4-105-86	A30	D30	108308	108326	E30	108328	108346
4-45-49	A31	D31	108453	108471	E31	108473	108491
4-44-277	A32	D32	109177	109195	E32	109197	109215
4-86-60	A33	D33	114585	114603	E33	114605	114623
4-84-334	A34	D34	115697	115715	E34	115717	115735
99-78-321	A35	D35	122064	122082	E35	122084	122102
99-12767-36	A36	D36	123105	123123	E36	123125	123143
99-12767-143	A37	D37	123212	123230	E37	123232	123250
99-12767-189	A38	D38	123258	123276	E38	123278	123296
99-12767-380	A39	D39	123449	123467	E39	123469	123487
4-80-328	A40	D40	126719	126737	E40	126739	126757
4-36-384	A41	D41	128191	128209	E41	128211	128229
4-36-264	A42	D42	128311	128329	E42	128331	128349
4-36-261	A43	D43	128314	128332	E43	128334	128352
4-35-333	A44	D44	128575	128593	E44	128595	128613
4-35-240	A45	D45	128668	128686	E45	128688	128706
4-35-173	A46	D46	128735	128753	E46	128755	128773
4-35-133	A47	D47	128775	128793	E47	128795	128813
99-12771-59	A48	D48	130786	130804	E48	130806	130824
99-12774-334	A49	D49	133187	133205	E49	133207	133225
99-12776-358	A50	D50	135367	135385	E50	135387	135405
99-12781-113	A51	D51	139370	139388	E51	139390	139408
4-104-298	A52	D52	157516	157534	E52	157536	157554
4-104-254	A53	D53	157560	157578	E53	157580	157598
4-104-250	A54	D54	157564	157582	E54	157584	157602
4-104-214	A55	D55	157600	157618	E55	157620	157638
99-12818-289	A56	D56	172961	172979	E56	172981	172999
99-24807-271	A57	D57	180603	180621	E57	180623	180641
99-24807-84	A58	D58	180790	180808	E58	180810	180828
99-12831-157	A59	D59	190315	190333	E59	190335	190353
99-12831-241	A60	D60	190399	190417	E60	190419	190437
99-12832-387	A61	D61	191378	191396	E61	191398	191416
99-12836-30	A62	D62	195109	195127	E62	195129	195147
99-12844-262	A63	D63	203827	203845	E63	203847	203865
4-24-74	A64	D64	210132	210150	E64	210152	210170
4-24-246	A65	D65	210302	210320	E65	210322	210340
4-24-314	A66	D66	210370	210388	E66	210390	210408
4-27-190	A67	D67	211149	211167	E67	211169	211187
5-400-145	A68	D68	215977	215995	E68	215997	216015
5-400-149	A69	D69	215981	215999	E69	216001	216019
5-400-175	A70	D70	216007	216025	E70	216027	216045
5-400-231	A71	D71	216063	216081	E71	216083	216101
5-400-367	A72	D72	216199	216217	E72	216219	216237
99-12852-110	A73	D73	216303	216321	E73	216323	216341
99-12852-325	A74	D74	216518	216536	E74	216538	216556
4-37-326	A75	D75	221630	221648	E75	221650	221668

4-37-107	A76	D76	221848	221866	E76	221868	221886
5-270-92	A77	D77	225626	225644	E77	225646	225664
99-12860-47	A78	D78	229368	229386	E78	229388	229406
99-12860-57	A79	D79	229378	229396	E79	229398	229416
5-402-144	A80	D80	237536	237554	E80	237556	237574

Mis 1 and Mis 2 respectively refer to microsequencing primers which hybridized with the non-coding strand of the PG-3 gene or with the coding strand of the PG-3 gene.

The microsequencing reaction was performed as follows :

After purification of the amplification products, the microsequencing reaction mixture was prepared by adding, in a 20µl final volume: 10 pmol microsequencing oligonucleotide, 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM $MgCl_2$), and the two appropriate fluorescent ddNTPs (Perkin Elmer, Dye Terminator Set 401095) complementary to the nucleotides at the polymorphic site of each biallelic marker tested, following the manufacturer's recommendations. After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The unincorporated dye terminators were then removed by ethanol precipitation. Samples were finally resuspended in formamide-EDTA loading buffer and heated for 2 min at 95°C before being loaded on a polyacrylamide sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Following gel analysis, data were automatically processed with software that allows the determination of the alleles of biallelic markers present in each amplified fragment.

The software evaluates such factors as whether the intensities of the signals resulting from the above microsequencing procedures are weak, normal, or saturated, or whether the signals are ambiguous. In addition, the software identifies significant peaks (according to shape and height criteria). Among the significant peaks, peaks corresponding to the targeted site are identified based on their position. When two significant peaks are detected for the same position, each sample is categorized classification as homozygous or heterozygous type based on the height ratio.

EXAMPLE 5

PREPARATION OF ANTIBODY COMPOSITIONS TO THE PG-3 PROTEIN

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the PG-3 protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the PG-3 protein or a portion thereof can be prepared fr m murine hybridomas according to the classical method of Kohler, G. and Milstein, C., (1975) or derivative methods thereof. Also see Harlow, E., and D. Lane. 1988.

Briefly, a mouse is repetitively inoculated with a few micrograms of the PG-3 protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* (1986).

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the PG-3 protein or a portion thereof can be prepared by immunizing suitable non-human animal with the PG-3 protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for PG-3 concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.* (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein by the one skilled in the art without departing from the spirit and scope of the invention.

REFERENCES

Abbondanzo SJ *et al.*, 1993, Methods in Enzymology, Academic Press, New York, pp 803-823

Ajioka R.S. *et al.*, *Am. J. Hum. Genet.*, 60:1439-1447, 1997

Altschul *et al.*, 1990, J. Mol. Biol. 215(3):403-410

Altschul *et al.*, 1993, Nature Genetics 3:266-272

Altschul *et al.*, 1997, Nuc. Acids Res. 25:3389-3402

Anton M. *et al.*, 1995, J. Virol., **69** : 4600-4606

Araki K *et al.* (1995) *Proc. Natl. Acad. Sci. U S A.* 92(1):160-4.

Arnheim N & Shibata D, Curr. Op. Genetics & Development, 1997, 7:364-370

Aszódi *et al.*, Proteins:Structure, Function, and Genetics, Supplement 1:38-42 (1997)

Ausubel *et al.* (1989)Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.

Baubonis W. (1993) *Nucleic Acids Res.* 21(9):2025-9.

Beaucage *et al.*, *Tetrahedron Lett* 1981, **22**: 1859-1862

Bochar *et al.*, (2000) *Cell* **102**:257-265

Bradley A., 1987, Production and analysis of chimaeric mice. *In*: E.J. Robertson (Ed.), Teratocarcinomas and embryonic stem cells: A practical approach. IRL Press, Oxford, pp.113.

Bram RJ *et al.*, 1993, Mol. Cell Biol., **13** : 4760-4769

Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979;**68**:109-151

Brutlag *et al.* Comp. App. Biosci. 6:237-245, 1990

Bush *et al.*, 1997, J. Chromatogr., **777** : 311-328.

Chai H. *et al.* (1993) *Biotechnol. Appl. Biochem.*18:259-273.

Chee *et al.* (1996) *Science.* 274:610-614.

Chen and Kwok *Nucleic Acids Research* 25:347-353 1997

Chen *et al.* (1987) *Mol. Cell. Biol.* 7:2745-2752.

Chen *et al. Proc. Natl. Acad. Sci. USA* 94/20 10756-10761,1997

Cho RJ *et al.*, 1998, Proc. Natl. Acad. Sci. USA, **95**(7) : 3752-3757.

Chou J.Y., 1989, Mol. Endocrinol., **3**: 1511-1514.

Clark A.G. (1990) *Mol. Biol. Evol.* 7:111-122.

Coles R, Caswell R, Rubinsztein DC, *Hum Mol Genet* 1998;7:791-800

Compton J. (1991) *Nature.* 350(6313):91-92.

Davis L.G., M.D. Dibner, and J.F. Battey, Basic Methods in Molecular Biology, ed., Elsevier Press, NY, 1986

Dempster *et al.*, (1977) *J. R. Stat. Soc.*, 39B:1-38.

Dent DS & Latchman DS (1993) The DNA mobility shift assay. In: *Transcription Factors: A Practical Approach* (Latchman DS, ed.) pp1-26. Oxford: IRL Press

Eckner R. *et al.* (1991) *EMBO J.* 10:3513-3522.

Edwards et Leatherbarrow, *Analytical Biochemistry*, **246**, 1-6 (1997)

Ellis NA,1997, Curr.Op.Genet.Dev.7:.354-363

Emi M, *et al.*, Cancer Res. 1992 Oct 1; 52(19): 5368-5372

Engvall, E., Meth. Enzymol. 70:419 (1980)

Excoffier L. and Slatkin M. (1995) *Mol. Biol. Evol.*, 12(5): 921-927.

Fanger GR *et al.*, 1997 Curr.Op.Genet.Dev.7:67-74

Feldman and Steg, 1996, Medecine/Sciences, synthese, 12:47-55

Felici F., 1991, J. Mol. Biol., Vol. 222:301-310

Fields and Song, 1989, Nature, **340** : 245-246

Fishel R & Wilson T. 1997, Curr.Op.Genet.Dev.7: 105-113;

Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980)

Flotte *et al.* (1992) *Am. J. Respir. Cell Mol. Biol.* 7:349-356.

Fodor *et al.* (1991) *Science* 251:767-777.

Fraley *et al.* (1979) *Proc. Natl. Acad. Sci. USA.* 76:3348-3352.

Fried M, Crothers DM, *Nucleic Acids Res* 1981;9:6505-6525

Fromont-Racine M. *et al.*, 1997, Nature Genetics, **16**(3) : 277-282.

Fuller S. A. *et al.* (1996) *Immunology in Current Protocols in Molecular Biology*, Ausubel *et al.*Eds, John Wiley & Sons, Inc., USA.

Furth P.A. *et al.* (1994) *Proc. Natl. Acad. Sci USA.* 91:9302-9306.

Garner MM, Revzin A, *Nucleic Acids Res* 1981;9:3047-3060

Geysen H. Mario *et al.* 1984. Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002

Ghosh and Bacchawat, 1991, *Targeting of liposomes to hepatocytes*, IN: *Liver Diseases, Targeted diagnosis and therapy using specific rceptors and ligands*. Wu *et al.* Eds., Marcel Dekeker, New York, pp. 87-104.

Gonnet *et al.*, 1992, Science 256:1443-1445

Gopal (1985) *Mol. Cell. Biol.*, 5:1188-1190.

Gossen M. *et al.* (1992) *Proc. Natl. Acad. Sci. USA.* 89:5547-5551.

Gossen M. *et al.* (1995) *Science.* 268:1766-1769.

Graham *et al.* (1973) *Virology* 52:456-457.

Green *et al.*, *Ann. Rev. Biochem.* 55:569-597 (1986)

Griffin *et al. Science* 245:967-971 (1989)

Grompe, M. (1993) *Nature Genetics.* 5:111-117.

Grompe, M. *et al.* (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:5855-5892.

Gronwald J, *et al.*, Cancer Res. 1997 Feb 1; 57(3): 481-487

Gu H. *et al.* (1993) *Cell* 73:1155-1164.

Gu H. *et al.* (1994) *Science* 265:103-106.

Guatelli J C *et al.* (1990) *Proc. Natl. Acad. Sci. USA.* 35:273-286.

Haber D & Harlow E, 1997, Nature Genet. 16:320-322

Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS, *Nat Genet* 1996;14(4):441-447

Haff L. A. and Smirnov I. P. (1997) *Genome Research*, 7:378-388.

Hames B.D. and Higgins S.J. (1985*) Nucleic Acid Hybridization: A Practical Approach.* Hames and Higgins Ed., IRL Press, Oxford.

Harju L, Weber T, Alexandrova L, Lukin M, Ranki M, Jalanko A, *Clin Chem* 1993;39(11Pt 1):2282-2287

Harland *et al.* (1985) *J. Cell. Biol.* 101:1094-1095.

Harlow, E., and D. Lane. 1988. Antibodies A Laboratory Manual. Cold Spring Harbor Laboratory. pp. 53-242

Harper JW *et al.*, 1993, Cell, 75 : 805-816

Harris H *et al.*,1969,Nature 223:363-368

Hawley M.E. *et al.* (1994) *Am. J. Phys. Anthropol.* 18:104.

Henikoff and Henikoff, 1993, Proteins 17:49-61

Higgins *et al.*, 1996, Methods Enzymol. 266:383-402

Hillier L. and Green P. *Methods Appl.*, 1991, 1: 124-8.

Hoess *et al.* (1986) *Nucleic Acids Res.* 14:2287-2300.

Huang L. *et al.* (1996) *Cancer Res* 56(5):1137-1141.

Hunter T, 1991 Cell 64:249

Huygen *et al.* (1996) *Nature Medicine.* 2(8):893-898.

Ichikawa T, *et al.*, Prostate Suppl. 1996; 6: 31-35

Ishwad CS, *et al.*, Int. J. Cancer. 1999 Jan 5; 80(1): 25-31

Izant JG, Weintraub H, *Cell* 1984 Apr;36(4):1007-15

Julan *et al.* (1992) *J. Gen. Virol.* 73:3251-3255.

Kanegae Y. *et al.*, *Nucl. Acids Res.* 23:3816-3821 (1995).

Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268

Khoury J. *et al., Fundamentals of Genetic Epidemiology, Oxford University Press, NY*, 1993

Kim U-J. *et al.* (1996) *Genomics* 34:213-218.

Klein *et al.* (1987) *Nature.* 327:70-73.

Kohler, G. and Milstein, C., Nature 256:495 (1975)

Koller *et al.* Proc. Natl. Acad. Sci. USA 86:8932-8935 (1989)

Koller *et al.* (1992) *Annu. Rev. Immunol.* 10:705-730.

Kozal MJ, Shah N, Shen N, Yang R, Fucini R, Merigan TC, Richman DD, Morris D, Hubbell E, Chee M, Gingeras TR, Nat Med 1996;2(7):753-759

Landegren U. *et al.* (1998) *Genome Research*, 8:769-776.

Lander and Schork, *Science,* 265, 2037-2048, 1994

Lange K. (1997) *Mathematical and Statistical Methods for Genetic Analysis*. Springer, New York.

Lenhard T. *et al.* (1996) *Gene.* 169:187-190.

Linton M.F. *et al.* (1993) *J. Clin. Invest.* 92:3029-3037.

Liu Z. *et al.* (1994) *Proc. Natl. Acad. Sci. USA.* 91: 4528-4262.

Livak *et al.*, *Nature Genetics*, 9:341-342, 1995

Livak KJ, Hainer JW, *Hum Mutat* 1994;3(4):379-385

Lockhart *et al. Nature Biotechnology* 14: 1675-1680, 1996

Lucas A.H., 1994, In : Development and Clinical Uses of Haempophilus b Conjugate;

Mansour S.L. *et al.* (1988) *Nature.* 336:348-352.

Marshall R. L. *et al.* (1994) *PCR Methods and Applications.* 4:80-84.

Matsuyama H, *et al.*, Oncogene 1994 Oct; 9(10): 3071-3076

McCormick *et al.* (1994) *Genet. Anal. Tech. Appl.* 11:158-164.

McLaughlin B.A. *et al.* (1996) *Am. J. Hum. Genet.* 59:561-569.

Morton N.E., *Am.J. Hum.Genet.*, 7:277-318, 1955

Muzyczka *et al.* (1992) *Curr. Topics in Micro. and Immunol.* 158:97-129.

Nada S. *et al.* (1993) *Cell* 73:1125-1135.

Nagai H, *et al.*, Oncogene 1997 Jun 19; 14(24): 2927-2933

Nagy A. *et al.*, 1993, Proc. Natl. Acad. Sci. USA, 90: 8424-8428.

Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;68:90-98

Neda *et al.* (1991) *J. Biol. Chem.* 266:14143-14146.

Newton *et al.* (1989) *Nucleic Acids Res.* 17:2503-2516.

Nickerson D.A. *et al.* (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:8923-8927.

Nicolau C. *et al.*, 1987, Methods Enzymol., **149**:157-76.

Nicolau *et al.* (1982) *Biochim. Biophys. Acta.* 721:185-190.

Nyren P, Pettersson B, Uhlen M, *Anal Biochem* 1993;**208**(1):171-175

O'Reilly *et al.* (1992) *Baculovirus Expression Vectors: A Laboratory Manual.* W. H. Freeman and Co., New York.

Ohno *et al.* (1994) *Science.* 265:781-784.

Oldenburg K.R. *et al.*, 1992, Proc. Natl. Acad. Sci., **89**:5393-5397.

Orita *et al.* (1989) *Proc. Natl. Acad. Sci. U.S.A.*86: 2776-2770.

Ott J., *Analysis of Human Genetic Linkage, John Hopkins University Press, Baltimore,* 1991

Ouchterlony, O. *et al.*, Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973)

Parmley and Smith, Gene, 1988, **73**:305-318

Pastinen *et al.*, *Genome Research* 1997; 7:606-614

Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85(8):2444-2448

Pease S. ans William R.S., 1990, Exp. Cell. Res., **190**: 209-211.

Perinchery G, *et al.*, Int. J. Oncol. 1999 Mar; 14(3): 495-500

Perlin *et al.* (1994) *Am. J. Hum. Genet.* 55:777-787.

Peterson *et al.*, 1993, Proc. Natl. Acad. Sci. USA, **90** : 7593-7597.

Pietu *et al. Genome Research* 6:492-503, 1996

Pineau P, *et al.*, Oncogene 1999 May 20; 18(20): 3127-3134

Potter *et al.* (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81(22):7161-7165.

Ramunsen *et al.*, 1997, Electrophoresis, **18** : 588-598.

Reid L.H. *et al.* (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:4299-4303.

Risch, N. and Merikangas, K. (*Science*, 273:1516-1517, 1996

Robertson E., 1987, Embryo-derived stem cell lines. In: E.J. Robertson Ed. *Teratocarcinomas and embrionic stem cells: a practical approach.* IRL Press, Oxford, pp. 71.

Rossi *et al.*, *Pharmacol. Ther.* **50**:245-254, (1991)

Roth J.A. *et al.* (1996) *Nature Medicine.* 2(9):985-991.

Roux *et al.* (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:9079-9083.

Ruano *et al.* (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:6296-6300.

Sakabe T, *et al.*, Cancer Res. 1999 Feb 1; 59(3): 511-515

Sakakura C, *et al.*, Genes Chromosomes Cancer 1999 Apr; 24(4): 299-305

Sambrook, J., Fritsch, E.F., and T. Maniatis. (1989) *Molecular Cloning: A Laboratory Manual.* 2ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Samson M, *et al.* (1996) *Nature*, 382(6593):722-725.

Samulski *et al.* (1989) *J. Virol.* 63:3822-3828.

Sanchez-Pescador R. (1988) *J. Clin. Microbiol.* 26(10):1934-1938.

Sarkar, G. and Sommer S.S. (1991) *Biotechniques*.

Sauer B. *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:5166-5170.

Schaid D.J. *et al.*, *Genet. Epidemiol.*,13:423-450, 1996

Schedl A. *et al.*, 1993a, Nature, **362**: 258-261.

Schedl *et al.*, 1993b, Nucleic Acids Res., **21**: 4783-4787.

Schena *et al. Science* **270**:467-470, 1995

Schena *et al.*, 1996, Proc Natl Acad Sci U S A,.**93**(20):10614-10619.

Schneider *et al.*(1997) *Arlequin: A Software For Population Genetics Data Analysis.* University of Geneva.

Scholnick SB, *et al.*, J. Natl. Cancer Inst. 1996 Nov 20; 88(22): 1676-1682

Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical Research Foundation

Sczakiel G. *et al.* (1995) *Trends Microbiol.* 3(6):213-217.

Shay J.W. *et al.*, 1991, Biochem. Biophys. Acta, **1072**: 1-7.

Sheffield, V.C. *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 49:699-706.

Shizuya *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:8794-8797.

Shoemaker DD, *et al.*, *Nat Genet* 1996;**14**(4):450-456

Smith (1957) *Ann. Hum. Genet.* 21:254-276.

Smith *et al.* (1983) *Mol. Cell. Biol.* 3:2156-2165.

Sosnowski RG, *et al.*, *Proc Natl Acad Sci U S A* 1997;**94**:1119-1123

Sowdhamini *et al.*, Protein Engineering 10:207, 215 (1997)

Spielmann S. and Ewens W.J., *Am. J. Hum. Genet.*, 62:450-458, 1998

Spielmann S. *et al.*, *Am. J. Hum. Genet.*, 52:506-516, 1993

Sternberg N.L. (1994) *Mamm. Genome.* 5:397-404.

Sternberg N.L. (1992) Trends Genet. 8:1-16.

Stryer, L., *Biochemistry,* 4th edition, 1995, W. H Freeman & Co., New York.

Sunwoo JB, *et al.*, Genes Chromosomes Cancer 1996 Jul; 16(3):164-169

Sunwoo JB, *et al.*, Oncogene 1999 Apr 22; 18(16): 2651-2655

Syvanen AC, *Clin Chim Acta* 1994;**226**(2):225-236

Szabo A. *et al. Curr Opin Struct Biol* **5**, 699-705 (1995)

Tacson *et al.* (1996) *Nature Medicine.* 2(8):888-892.

Te Riele *et al.* (1990) Nature. 348:649-651.

Terwilliger J.D. and Ott J., *Handbook of Human Genetic Linkage, John Hopkins University Press, London*, 1994

Thomas K.R. *et al.* (1986) *Cell.* 44:419-428.

Thomas K.R. *et al.* (1987) *Cell.* 51:503-512.

Thompson *et al.*, 1994, Nucleic Acids Res. 22(2):4673-4680

Tur-Kaspa *et al.* (1986) *Mol. Cell. Biol.* 6:716-718.

Tyagi *et al.* (1998) *Nature Biotechnology.* 16:49-53.

Urdea M.S. (1988) *Nucleic Acids Research.* 11:4937-4957.

Urdea M.S. *et al.*(1991) *Nucleic Acids Symp. Ser.* 24:197-200.

Vaitukaitis, J. *et al.* J. Clin. Endocrinol. Metab. 33:988-991 (1971)

Valadon P., *et al.*, 1996, J. Mol. Biol., **261**:11-22.

Van der Lugt *et al.* (1991) *Gene.* 105:263-267.

Vlasak R. *et al.* (1983) *Eur. J. Biochem.* 135:123-126.

Wabiko *et al.* (1986) *DNA.*5(4):305-314.

Walker *et al.* (1996) *Clin. Chem.* 42:9-13.

Wang *et al.*, 1997, Chromatographia, **44** : 205-208.

Washbum J, Woino K, and Macoska J, Proceedings of American Association for Cancer Research, March 1997; 38

Weir, B.S. (1996) *Genetic data Analysis II: Methods for Discrete population genetic Data, Sinauer Assoc., Inc., Sunderland, MA, U.S.A.*

Weiss FU *et al.*, 1997 Curr.Op.Genet.Dev.7:80-86

Westerink M.A.J., 1995, Proc. Natl. Acad. Sci., **92**:4021-4025

White, M.B. *et al.* (1992) *Genomics.* 12:301-306.

Wong *et al.* (1980) *Gene.* 10:87-94.

Wood S.A. *et al.*, 1993, Proc. Natl. Acad. Sci. USA, **90**: 4582-4585.

Wright K, *et al.*, Oncogene 1998 Sep 3; 17(9): 1185-1188

Wu and Wu (1987) *J. Biol. Chem.* 262:4429-4432.

Wu and Wu (1988) *Biochemistry.* 27:887-892.

Wu *et al.* (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:2757.

Yagi T. *et al.* (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:9918-9922.

Yaremko ML, *et al.*, Genes Chromosomes Cancer 1994 May;10(1):1-6

Zhao *et al.*, *Am. J. Hum. Genet.*, 63:225-240, 1998

Zou Y. R. *et al.* (1994) *Curr. Biol.* 4:1099-1103.

SEQUENCE LISTING FREE TEXT

The following free text appears in the accompanying Sequence Listing :

5' regulatory region

3' regulatory region

polymorphic base

or

complement

probe

sequencing oligonucleotide primer

insertion of

exon

WHAT IS CLAIMED:

1. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 15 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825.

2. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 15 nucleotides of SEQ ID No 2 or the complements thereof.

3. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of at least 15 nucleotides of anyone of SEQ ID Nos 1 and 2 or the complement thereof, wherein said span includes a PG-3-related biallelic marker in said sequence.

4. A polynucleotide according to claim 3, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof.

5. A polynucleotide according to claim 3, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof.

6. A polynucleotide according to claim 3, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof.

7. A polynucleotide according to claim 3, wherein said contiguous span is 18 to 35 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide.

8. A polynucleotide according to claim 7, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide.

9. A polynucleotide according to claim 7, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: P1 to P4 and P6 to P80, and the complementary sequences thereto.

10. A polynucleotide according to any one of claims 1, 2 or 3, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide.

11. A polynucleotide according to claim 3, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

12. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of at least 15 nucleotides of anyone of SEQ ID No 1,2 or the complements thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of a PG-3-related biallelic marker in said sequence.

13. A polynucleotide according to claim 12, wherein the 3' end of said polynucleotide is located one nucleotide upstream of said PG-3-related biallelic marker in said sequence.

14. A polynucleotide according to claim 13, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: D1 to D4, D6 to D80, E1 to E4, and E6 to E80.

15. An isolated, purified, or recombinant polynucleotide consisting essentially of a sequence selected from the following sequences: B1 to B52 and C1 to C52.

16. An isolated, purified, or recombinant polynucleotide which encodes a polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 3.

17. A polynucleotide according to any one of claims 1-16 attached to a solid support.

18. An array of polynucleotides comprising at least one polynucleotide according to claim 17.

19. An array according to claim 18, wherein said array is addressable.

20. A polynucleotide according to any one of claims 1-16 further comprising a label.

21. A recombinant vector comprising a polynucleotide according to any one of claims 1-16.

22. A host cell comprising a recombinant vector according to claim 21.

23. A non-human host animal or mammal comprising a recombinant vector according to claim 21.

24. A mammalian host cell comprising a PG-3 gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1- 16.

25. A non-human host mammal comprising a PG-3 gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1-16.

26. A method of genotyping comprising determining the identity of a nucleotide at a PG-3-related biallelic marker or the complement thereof in a biological sample.

27. A method according to claim 26, wherein said biological sample is derived from a single subject.

28. A method according to claim 27, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome.

29. A method according to claim 26, wherein said biological sample is derived from multiple subjects.

30. A method according to claim 26, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step.

31. A method according to claim 30, wherein said amplifying is performed by PCR.

32. A method according to claim 26, wherein said determining is performed by a hybridization assay.

33. A method according to claim 26, wherein said determining is performed by a sequencing assay.

34. A method according to claim 26, wherein said determining is performed by a microsequencing assay.

35. A method according to claim 26, wherein said determining is performed by an enzyme-based mismatch detection assay.

36. A method of estimating the frequency of an allele of a PG-3-related biallelic marker in a population comprising:

a) genotyping individuals from said population for said biallelic marker according to the method of claim 26; and

b) determining the proportional representation of said biallelic marker in said population.

37. A method of detecting an association between a genotype and a trait, comprising the steps of:

a) determining the frequency of at least one PG-3-related biallelic marker in trait positive population according to the method of claim 36;

b) determining the frequency of at least one PG-3-related biallelic marker in a control population according to the method of claim 36; and

c) determining whether a statistically significant association exists between said genotype and said trait.

38. A method of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising:

a) genotyping at least one PG-3-related biallelic marker according to claim 27 for each individual in said population;

b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome of each individual in said population; and

c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

39. A method according to claim 38, wherein said haplotype determination method is selected from the group consisting of asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark algorithm, or an expectation-maximization algorithm.

40. A method of detecting an association between a haplotype and a trait, comprising the steps of:

a) estimating the frequency of at least one haplotype in a trait positive population according to the method of claim 38;

b) estimating the frequency of said haplotype in a control population according to the method of claim 38; and

c) determining whether a statistically significant association exists between said haplotype and said trait.

41. A method according to claim 37, wherein said genotyping steps a) and b) are performed on a single pooled biological sample derived from each of said populations.

42. A method according to claim 37, wherein said genotyping steps a) and b) performed separately on biological samples derived from each individual in said populations.

43. A method according to either claim 37 or 40, wherein said trait is cancer susceptibility.

44. A method according to either claim 37 or 40, wherein said control population is a trait negative population.

45. A method according to either claim 37 or 40, wherein said case control population is a random population.

46. Use of a polynucleotide comprising a contiguous span of at least 15 nucleotides of a sequence selected from the group consisting of the SEQ ID Nos 1, 2, amplicons 5-390, 5-391, 5-392, 4-59, 4-58, 4-54, 4-51, 99-86, 4-88, 5-397, 5-398, 99-12738, 99-109, 99-12749, 4-21, 4-23, 99-12753, 5-364, 99-12755, 4-87, 99-12757, 99-12758, 4-105, 4-45, 4-44, 4-86, 4-84, 99-78, 99-12767, 4-80, 4-36, 4-35, 99-12771, 99-12774, 99-12776, 99-12781, 4-104, 99-12818, 99-24807, 99-12827, 99-12831, 99-12832, 99-12836, 99-12844, 4-24, 4-27, 5-400, 99-12852, 4-37, 5-270, 99-12860, and 5-402 or the complementary sequence thereto for determining the identity of the nucleotide at a PG-3-related biallelic marker

47. Use according to claim 46 in a microsequencing assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said PG-3-related biallelic marker in said sequence.

48. Use according to claim 46 in a hybridization assay, wherein said c ntiguous span includes said PG-3-related biallelic marker.

49. Use according to claim 46 in a specific amplification assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

50. Use according to claim 46 in a sequencing assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide.

51. Use according to any one of claims 46-50, wherein said PG-3-related biallelic is a biallelic marker selected from the group consisting of A1 to A80.

52. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 3.

53. An isolated or purified antibody composition capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 52.

54. A method according to any one of claims 26-45, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A80 and the complements thereof.

55. A diagnostic kit comprising a polynucleotide according to any one of claims 3-15.

56. A computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 15 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825;

b) a contiguous span of at least 15 nucleotides of SEQ ID No 2 or the complements thereof; and

c) a nucleotide sequence complementary to any one of the preceding nucleotide sequences.

57. A computer readable medium having stored thereon a sequence consisting of a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 3.

58. A computer system comprising a processor and a data storage device wherein said data storage device is a computer readable medium according to claim 56 or 57.

59. A computer system according to claim 58, further comprising a sequence comparer and a data storage device having reference sequences stored thereon.

60. A computer system of Claim 59 wherein said sequence comparer comprises a computer program which indicates polymorphisms.

61. A computer system of Claim 58 further comprising an identifier which identifies features in said sequence.

62. A method for comparing a first sequence to a reference sequence, comprising the steps of:

reading said first sequence and said reference sequence through use of a computer program which compares sequences; and

determining differences between said first sequence and said reference sequence with said computer program,

wherein said first sequence is selected from the group consisting of a nucleic acid code comprising one of the following:

a) a contiguous span of at least 15 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825;

b) a contiguous span of at least 15 nucleotides of SEQ ID No 2 or the complements thereof; and

c) a nucleotide sequence complementary to any one of the preceding nucleotide sequences; and,

d) a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 3.

AMENDED CLAIMS

[received by the International Bureau on 25 January 2001 (25.01.01); original claims 1,2 and 56 amended; remaining claims unchanged (2 pages)]

1. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 200 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825.

2. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 200 nucleotides of SEQ ID No 2 or the complements thereof.

3. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of at least 15 nucleotides of anyone of SEQ ID Nos 1 and 2 or the complement thereof, wherein said span includes a PG-3-related biallelic marker in said sequence.

4. A polynucleotide according to claim 3, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A80, and the complements thereof.

5. A polynucleotide according to claim 3, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A5 and A8 to A80, and the complements thereof.

6. A polynucleotide according to claim 3, wherein said PG-3-related biallelic marker is selected from the group consisting of A6 and A7, and the complements thereof.

7. A polynucleotide according to claim 3, wherein said contiguous span is 18 to 35 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide.

8. A polynucleotide according to claim 7, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide.

9. A polynucleotide according to claim 7, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: P1 to P4 and P6 to P80, and the complementary sequences thereto.

10. A polynucleotide according to any one of claims 1, 2 or 3, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide.

49. Use according to claim 46 in a specific amplification assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.

50. Use according to claim 46 in a sequencing assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide.

51. Use according to any one of claims 46-50, wherein said PG-3-related biallelic is a biallelic marker selected from the group consisting of A1 to A80.

52. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 6 amino acids of SEQ ID No 3.

53. An isolated or purified antibody composition capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 52.

54. A method according to any one of claims 26-45, wherein said PG-3-related biallelic marker is selected from the group consisting of A1 to A80 and the complements thereof.

55. A diagnostic kit comprising a polynucleotide according to any one of claims 3-15.

56. A computer readable medium having stored thereon at least 2 nucleic acid code sequences comprising any one of the following:

a) a contiguous span of at least 200 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 1-97921, 98517-103471, 103603-108222, 108390-109221, 109324-114409, 114538-115723, 115957-122102, 122225-126876, 127033-157212, 157808-240825;

b) a contiguous span of at least 200 nucleotides of SEQ ID No 2 or the complements thereof; and

c) a nucleotide sequence complementary to any one of the preceding nucleotide sequences.

57. A computer readable medium having stored thereon a sequence consisting of a polypeptide code comprising a contiguous span of at least 6 amino acids of SEQ ID No 3.

58. A computer system comprising a processor and a data storage device wherein said data storage device is a computer readable medium according to claim 56 or 57.

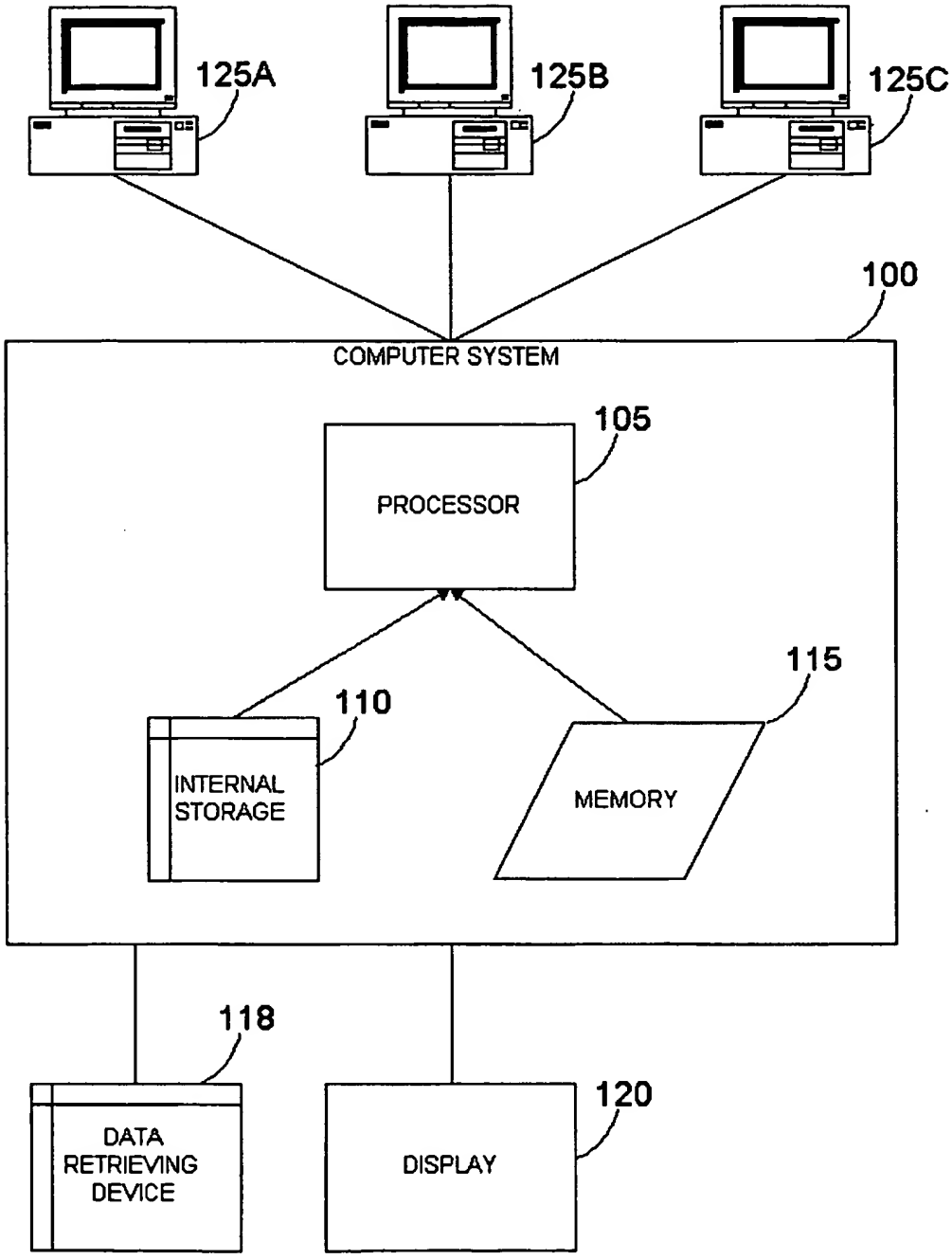
125A
125B
125C
100
COMPUTER SYSTEM
105
PROCESSOR
115
110
INTERNAL STORAGE
MEMORY
118
DATA RETRIEVING DEVICE
120
DISPLAY

Figure 1

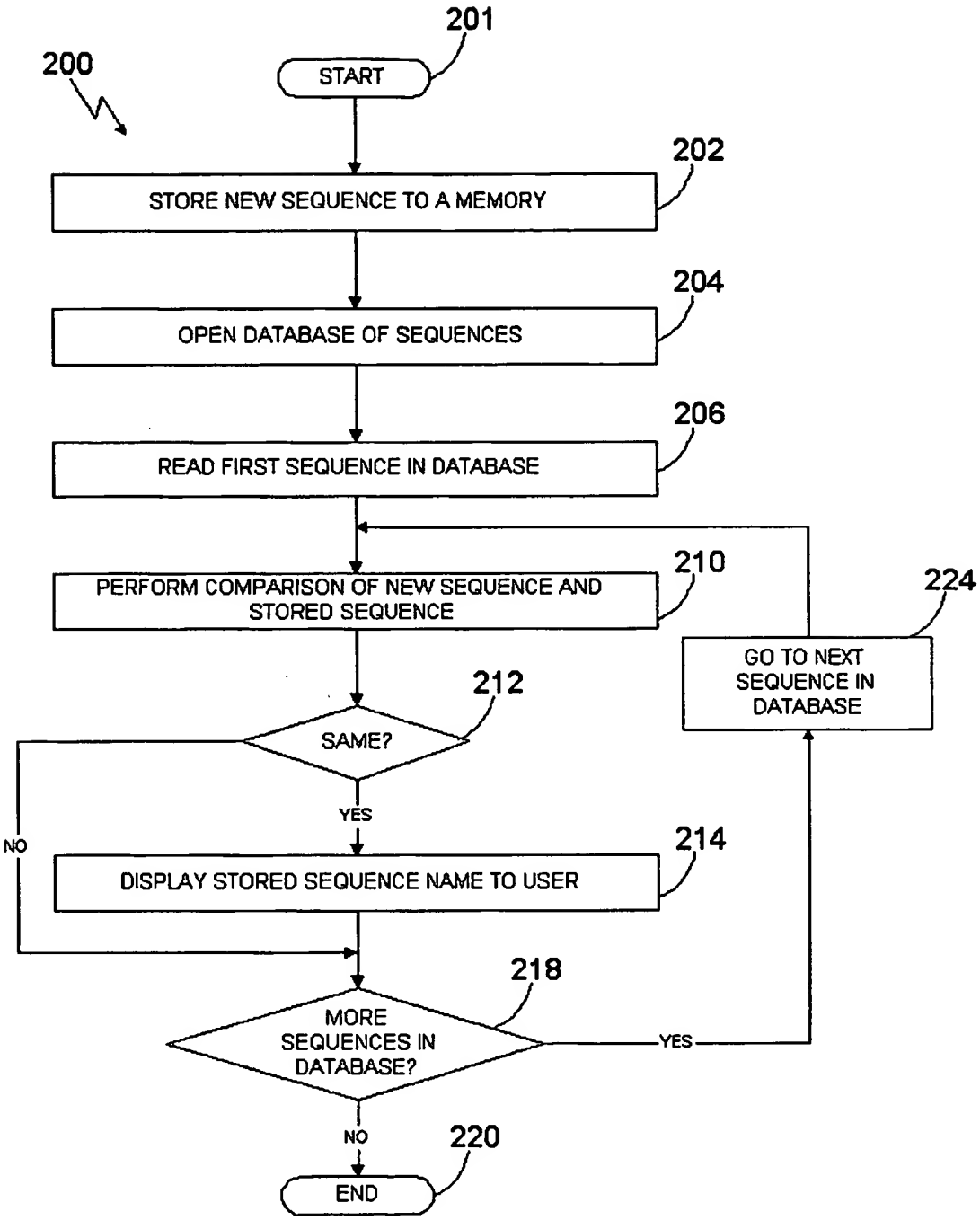
200
201
START
202
STORE NEW SEQUENCE TO A MEMORY
204
OPEN DATABASE OF SEQUENCES
206
READ FIRST SEQUENCE IN DATABASE
210
PERFORM COMPARISON OF NEW SEQUENCE AND STORED SEQUENCE
212
SAME?
YES
NO
214
DISPLAY STORED SEQUENCE NAME TO USER
218
MORE SEQUENCES IN DATABASE?
YES
NO
220
END
224
GO TO NEXT SEQUENCE IN DATABASE

Figure 2

<221> primer_bind
<222> 2108..2125
<223> 5-390.rp complement

<220>
<221> primer_bind
<222> 4559..4577
<223> 5-391.pu

<220>
<221> primer_bind
<222> 4891..4908
<223> 5-391.rp complement

<220>
<221> primer_bind
<222> 10007..10025
<223> 5-392.pu

<220>
<221> primer_bind
<222> 10411..10430
<223> 5-392.rp complement

<220>
<221> primer_bind
<222> 39556..39574
<223> 4-59.rp

<220>
<221> primer_bind
<222> 39877..39896
<223> 4-58.rp

<220>
<221> primer_bind
<222> 39953..39970
<223> 4-59.pu complement

<220>
<221> primer_bind
<222> 40242..40259
<223> 4-58.pu complement

<220>
<221> primer_bind
<222> 41137..41154
<223> 4-54.rp

<220>
<221> primer_bind
<222> 41564..41581
<223> 4-54.pu complement

<220>
<221> primer_bind
<222> 42122..42141
<223> 4-51.rp

<220>
<221> primer_bind
<222> 42526..42543
<223> 4-51.pu complement

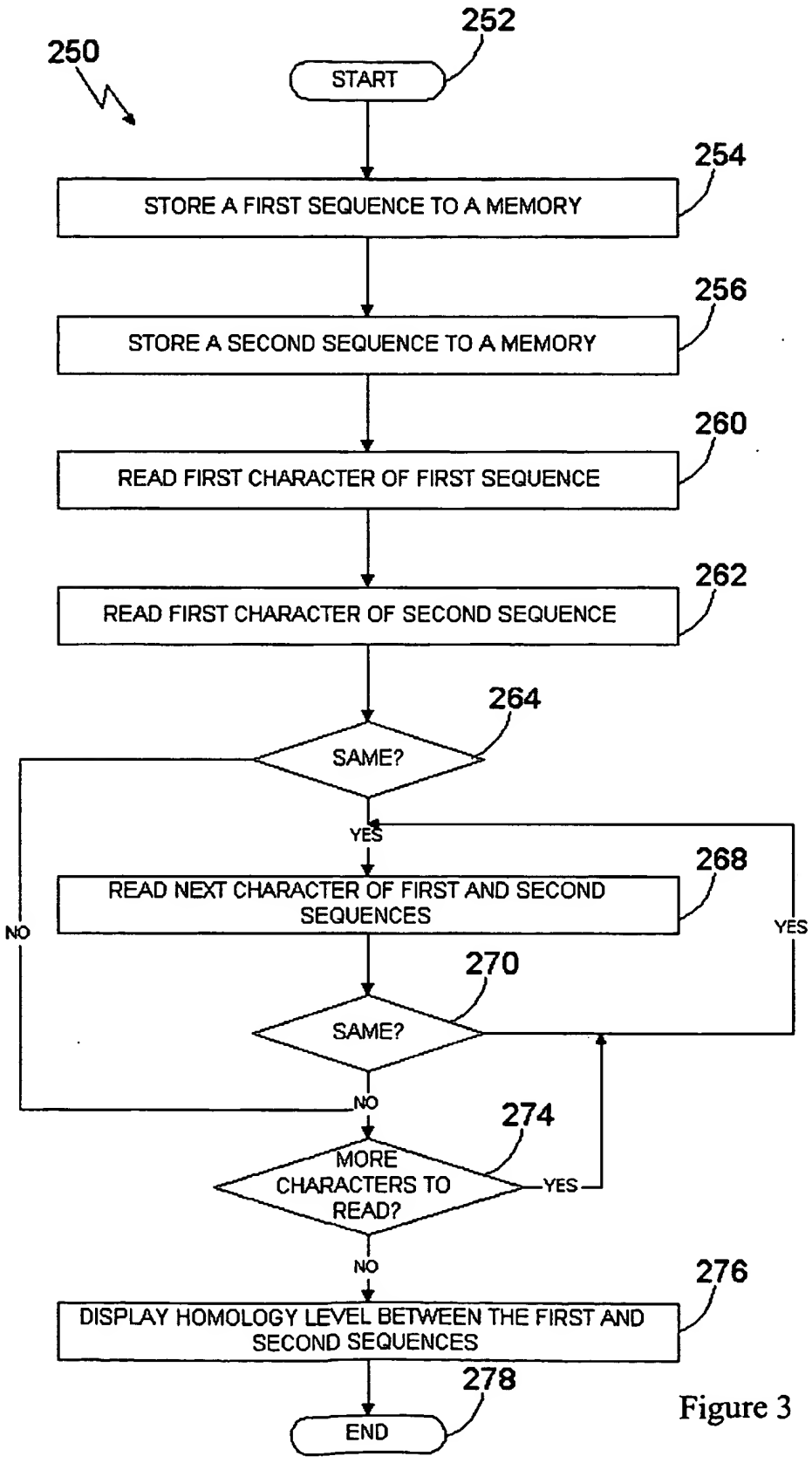
250
252
START
254
STORE A FIRST SEQUENCE TO A MEMORY
256
STORE A SECOND SEQUENCE TO A MEMORY
260
READ FIRST CHARACTER OF FIRST SEQUENCE
262
READ FIRST CHARACTER OF SECOND SEQUENCE
264
SAME?
YES
NO
268
READ NEXT CHARACTER OF FIRST AND SECOND SEQUENCES
YES
270
SAME?
NO
274
MORE CHARACTERS TO READ?
YES
NO
276
DISPLAY HOMOLOGY LEVEL BETWEEN THE FIRST AND SECOND SEQUENCES
278
END

Figure 3

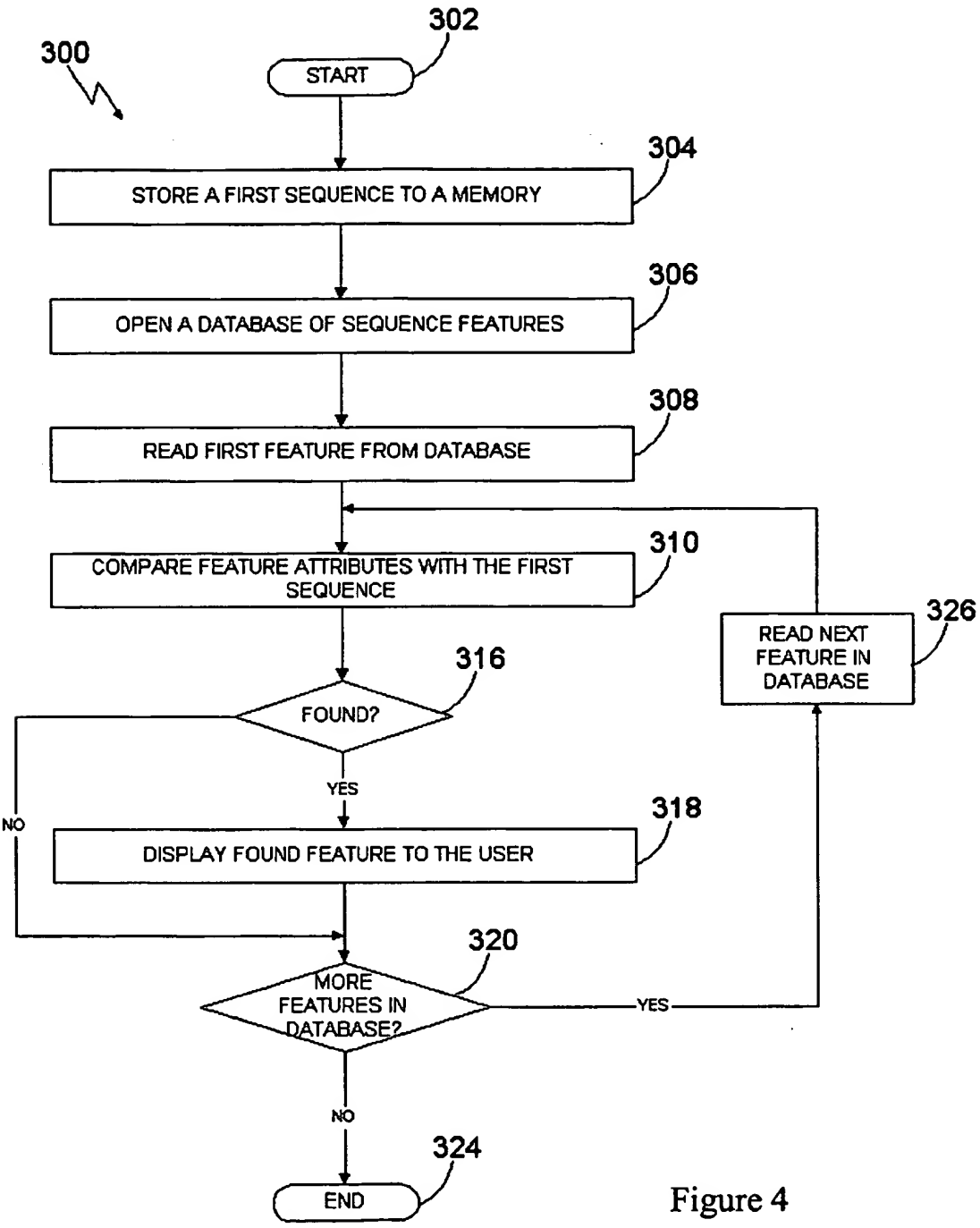
300
302
START
304
STORE A FIRST SEQUENCE TO A MEMORY
306
OPEN A DATABASE OF SEQUENCE FEATURES
308
READ FIRST FEATURE FROM DATABASE
310
COMPARE FEATURE ATTRIBUTES WITH THE FIRST SEQUENCE
316
FOUND?
YES
NO
318
DISPLAY FOUND FEATURE TO THE USER
320
MORE FEATURES IN DATABASE?
YES
NO
324
END
326
READ NEXT FEATURE IN DATABASE

Figure 4

<110> Genset

<120> PG-3 and biallelic markers thereof

<130> 68.WO1

<140> US 60/149,941
<141> 1999-08-19

<160> 5

<170> Patent.pm

<210> 1
<211> 240825
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 1..2000
<223> 5'regulatory region

<220>
<221> exon
<222> 2001..2079
<223> exon A

<220>
<221> exon
<222> 4627..4718
<223> exon B

<220>
<221> exon
<222> 10115..10233
<223> exon C

<220>
<221> exon
<222> 26810..26897
<223> exon D

<220>
<221> exon
<222> 31357..31471
<223> exon E

<220>
<221> exon
<222> 34261..34404
<223> exon F

<220>
<221> exon
<222> 37377..37466
<223> exon S

<220>
<221> exon
<222> 39704..40858
<223> exon T

<220>
<221> exon
<222> 50436..50545
<223> exon G

<220>
<221> exon
<222> 72881..72918
<223> exon H

<220>
<221> exon
<222> 75989..76151
<223> exon I

<220>
<221> exon
<222> 95111..95188
<223> exon J

<220>
<221> exon
<222> 216015..216252
<223> exon K

<220>
<221> exon
<222> 237526..238825
<223> exon L

<220>
<221> misc_feature
<222> 238826..240825
<223> 3'regulatory region

<220>
<221> allele
<222> 1999
<223> 5-390-177 : polymorphic base G or C

<220>
<221> allele
<222> 4601
<223> 5-391-43 : polymorphic base A or G

<220>
<221> allele
<222> 10228
<223> 5-392-222 : polymorphic base G or T

<220>
<221> allele
<222> 10286
<223> 5-392-280 : polymorphic base G or T

<220>
<221> allele
<222> 10370
<223> 5-392-364 : insertion of G

<220>
<221> allele
<222> 39944

<223> 4-58-318 : polymorphic base G or T

<220>
<221> allele
<222> 39973
<223> 4-58-289 : polymorphic base G or C

<220>
<221> allele
<222> 41385
<223> 4-54-199 : polymorphic base A or C

<220>
<221> allele
<222> 41404
<223> 4-54-180 : polymorphic base A or C

<220>
<221> allele
<222> 42232
<223> 4-51-312 : polymorphic base G or C

<220>
<221> allele
<222> 67475
<223> 99-86-266 : polymorphic base A or G

<220>
<221> allele
<222> 69521
<223> 4-88-107 : polymorphic base A or G

<220>
<221> allele
<222> 72838
<223> 5-397-141 : polymorphic base G or T

<220>
<221> allele
<222> 76060
<223> 5-398-203 : polymorphic base A or C

<220>
<221> allele
<222> 81253
<223> 99-12738-248 : polymorphic base A or C

<220>
<221> allele
<222> 83921
<223> 99-109-358 : polymorphic base A or C

<220>
<221> allele
<222> 91917
<223> 99-12749-175 : polymorphic base C or T

<220>
<221> allele
<222> 95349
<223> 4-21-154 : polymorphic base C or T

<220>

<221> allele
<222> 95511
<223> 4-21-317 : polymorphic base G or T

<220>
<221> allele
<222> 96190
<223> 4-23-326 : polymorphic base A or G

<220>
<221> allele
<222> 97294
<223> 99-12753-34 : polymorphic base A or T

<220>
<221> allele
<222> 98024
<223> 5-364-252 : polymorphic base G or T

<220>
<221> allele
<222> 98914
<223> 99-12755-280 : polymorphic base A or G

<220>
<221> allele
<222> 98963
<223> 99-12755-329 : polymorphic base A or C

<220>
<221> allele
<222> 103593
<223> 4-87-212 : polymorphic base A or G

<220>
<221> allele
<222> 104398
<223> 99-12757-318 : polymorphic base C or T

<220>
<221> allele
<222> 106373
<223> 99-12758-102 : polymorphic base A or G

<220>
<221> allele
<222> 106407
<223> 99-12758-136 : polymorphic base C or T

<220>
<221> allele
<222> 108315
<223> 4-105-98 : polymorphic base A or G

<220>
<221> allele
<222> 108327
<223> 4-105-86 : polymorphic base A or G

<220>
<221> allele
<222> 108472
<223> 4-45-49 : polymorphic base C or T

<220>
<221> allele
<222> 109196
<223> 4-44-277 : polymorphic base C or T

<220>
<221> allele
<222> 114604
<223> 4-86-60 : polymorphic base G or C

<220>
<221> allele
<222> 115716
<223> 4-84-334 : polymorphic base A or G

<220>
<221> allele
<222> 122083
<223> 99-78-321 : polymorphic base A or T

<220>
<221> allele
<222> 123124
<223> 99-12767-36 : polymorphic base G or C

<220>
<221> allele
<222> 123231
<223> 99-12767-143 : polymorphic base C or T

<220>
<221> allele
<222> 123277
<223> 99-12767-189 : polymorphic base C or T

<220>
<221> allele
<222> 123468
<223> 99-12767-380 : polymorphic base A or G

<220>
<221> allele
<222> 126738
<223> 4-80-328 : polymorphic base C or T

<220>
<221> allele
<222> 128210
<223> 4-36-384 : polymorphic base G or C

<220>
<221> allele
<222> 128330
<223> 4-36-264 : polymorphic base A or G

<220>
<221> allele
<222> 128333
<223> 4-36-261 : polymorphic base A or C

<220>
<221> allele

<222> 128594
<223> 4-35-333 : polymorphic base A or C

<220>
<221> allele
<222> 128687
<223> 4-35-240 : polymorphic base G or C

<220>
<221> allele
<222> 128754
<223> 4-35-173 : polymorphic base A or T

<220>
<221> allele
<222> 128794
<223> 4-35-133 : polymorphic base C or T

<220>
<221> allele
<222> 130805
<223> 99-12771-59 : polymorphic base G or T

<220>
<221> allele
<222> 133206
<223> 99-12774-334 : polymorphic base A or C

<220>
<221> allele
<222> 135386
<223> 99-12776-358 : polymorphic base A or G

<220>
<221> allele
<222> 139389
<223> 99-12781-113 : polymorphic base A or G

<220>
<221> allele
<222> 157535
<223> 4-104-298 : polymorphic base G or C

<220>
<221> allele
<222> 157579
<223> 4-104-254 : polymorphic base A or G

<220>
<221> allele
<222> 157583
<223> 4-104-250 : polymorphic base C or T

<220>
<221> allele
<222> 157619
<223> 4-104-214 : polymorphic base A or G

<220>
<221> allele
<222> 172980
<223> 99-12818-289 : polymorphic base C or T

<220>
<221> allele
<222> 180622
<223> 99-24807-271 : polymorphic base C or T

<220>
<221> allele
<222> 180809
<223> 99-24807-84 : polymorphic base A or G

<220>
<221> allele
<222> 190334
<223> 99-12831-157 : polymorphic base A or G

<220>
<221> allele
<222> 190418
<223> 99-12831-241 : polymorphic base C or T

<220>
<221> allele
<222> 191397
<223> 99-12832-387 : polymorphic base C or T

<220>
<221> allele
<222> 195128
<223> 99-12836-30 : polymorphic base G or C

<220>
<221> allele
<222> 203846
<223> 99-12844-262 : polymorphic base G or C

<220>
<221> allele
<222> 210151
<223> 4-24-74 : polymorphic base C or T

<220>
<221> allele
<222> 210321
<223> 4-24-246 : polymorphic base C or T

<220>
<221> allele
<222> 210389
<223> 4-24-314 : polymorphic base G or C

<220>
<221> allele
<222> 211168
<223> 4-27-190 : polymorphic base A or G

<220>
<221> allele
<222> 215996
<223> 5-400-145 : polymorphic base A or G

<220>
<221> allele
<222> 216000

<223> 5-400-149 : polymorphic base G or C

<220>
<221> allele
<222> 216026
<223> 5-400-175 : polymorphic base C or T

<220>
<221> allele
<222> 216082
<223> 5-400-231 : polymorphic base C or T

<220>
<221> allele
<222> 216218
<223> 5-400-367 : polymorphic base A or C

<220>
<221> allele
<222> 216322
<223> 99-12852-110 : polymorphic base G or T

<220>
<221> allele
<222> 216537
<223> 99-12852-325 : polymorphic base A or G

<220>
<221> allele
<222> 221649
<223> 4-37-326 : polymorphic base A or C

<220>
<221> allele
<222> 221867
<223> 4-37-107 : polymorphic base A or G

<220>
<221> allele
<222> 225645
<223> 5-270-92 : polymorphic base G or C

<220>
<221> allele
<222> 229387
<223> 99-12860-47 : polymorphic base A or G

<220>
<221> allele
<222> 229397
<223> 99-12860-57 : polymorphic base A or T

<220>
<221> allele
<222> 237555
<223> 5-402-144 : polymorphic base C or T

<220>
<221> primer_bind
<222> 1823..1840
<223> 5-390.pu

<220>

<220>
<221> primer_bind
<222> 67289..67309
<223> 99-86.rp

<220>
<221> primer_bind
<222> 67724..67741
<223> 99-86.pu complement

<220>
<221> primer_bind
<222> 69182..69200
<223> 4-88.rp

<220>
<221> primer_bind
<222> 69609..69626
<223> 4-88.pu complement

<220>
<221> primer_bind
<222> 72698..72715
<223> 5-397.pu

<220>
<221> primer_bind
<222> 73099..73117
<223> 5-397.rp complement

<220>
<221> primer_bind
<222> 75858..75877
<223> 5-398.pu

<220>
<221> primer_bind
<222> 76289..76306
<223> 5-398.rp complement

<220>
<221> primer_bind
<222> 81006..81025
<223> 99-12738.pu

<220>
<221> primer_bind
<222> 81466..81485
<223> 99-12738.rp complement

<220>
<221> primer_bind
<222> 83564..83582
<223> 99-109.pu

<220>
<221> primer_bind
<222> 83990..84007
<223> 99-109.rp complement

<220>
<221> primer_bind

<222> 91743..91763
<223> 99-12749.pu

<220>
<221> primer_bind
<222> 92123..92142
<223> 99-12749.rp complement

<220>
<221> primer_bind
<222> 95196..95214
<223> 4-21.pu

<220>
<221> primer_bind
<222> 95600..95619
<223> 4-21.rp complement

<220>
<221> primer_bind
<222> 95865..95882
<223> 4-23.pu

<220>
<221> primer_bind
<222> 96210..96229
<223> 4-23.rp complement

<220>
<221> primer_bind
<222> 97261..97278
<223> 99-12753.pu

<220>
<221> primer_bind
<222> 97728..97747
<223> 99-12753.rp complement

<220>
<221> primer_bind
<222> 97831..97849
<223> 5-364.rp

<220>
<221> primer_bind
<222> 98256..98275
<223> 5-364.pu complement

<220>
<221> primer_bind
<222> 98638..98656
<223> 99-12755.pu

<220>
<221> primer_bind
<222> 99111..99131
<223> 99-12755.rp complement

<220>
<221> primer_bind
<222> 103376..103395
<223> 4-87.rp

<220>
<221> primer_bind
<222> 103801..103818
<223> 4-87.pu complement

<220>
<221> primer_bind
<222> 104081..104100
<223> 99-12757.pu

<220>
<221> primer_bind
<222> 104619..104636
<223> 99-12757.rp complement

<220>
<221> primer_bind
<222> 106272..106291
<223> 99-12758.pu

<220>
<221> primer_bind
<222> 106780..106799
<223> 99-12758.rp complement

<220>
<221> primer_bind
<222> 108200..108218
<223> 4-105.rp

<220>
<221> primer_bind
<222> 108223..108246
<223> 4-45.rp

<220>
<221> primer_bind
<222> 108390..108412
<223> 4-105.pu complement

<220>
<221> primer_bind
<222> 108499..108520
<223> 4-45.pu complement

<220>
<221> primer_bind
<222> 109123..109142
<223> 4-44.rp

<220>
<221> primer_bind
<222> 109454..109471
<223> 4-44.pu complement

<220>
<221> primer_bind
<222> 114217..114234
<223> 4-86.rp

<220>
<221> primer_bind
<222> 114646..114663

<223> 4-86.pu complement

<220>
<221> primer_bind
<222> 115630..115647
<223> 4-84.rp

<220>
<221> primer_bind
<222> 116031..116049
<223> 4-84.pu complement

<220>
<221> primer_bind
<222> 121991..122011
<223> 99-78.rp

<220>
<221> primer_bind
<222> 122384..122401
<223> 99-78.pu complement

<220>
<221> primer_bind
<222> 123089..123106
<223> 99-12767.pu

<220>
<221> primer_bind
<222> 123565..123583
<223> 99-12767.rp complement

<220>
<221> primer_bind
<222> 126711..126729
<223> 4-80.rp

<220>
<221> primer_bind
<222> 127048..127065
<223> 4-80.pu complement

<220>
<221> primer_bind
<222> 128162..128179
<223> 4-36.rp

<220>
<221> primer_bind
<222> 128480..128497
<223> 4-35.rp

<220>
<221> primer_bind
<222> 128573..128590
<223> 4-36.pu complement

<220>
<221> primer_bind
<222> 128909..128926
<223> 4-35.pu complement

<220>

<221> primer_bind
<222> 130747..130764
<223> 99-12771.pu

<220>
<221> primer_bind
<222> 131254..131273
<223> 99-12771.rp complement

<220>
<221> primer_bind
<222> 132873..132892
<223> 99-12774.pu

<220>
<221> primer_bind
<222> 133305..133325
<223> 99-12774.rp complement

<220>
<221> primer_bind
<222> 135029..135048
<223> 99-12776.pu

<220>
<221> primer_bind
<222> 135458..135478
<223> 99-12776.rp complement

<220>
<221> primer_bind
<222> 139277..139296
<223> 99-12781.pu

<220>
<221> primer_bind
<222> 139724..139742
<223> 99-12781.rp complement

<220>
<221> primer_bind
<222> 157181..157199
<223> 4-104.rp

<220>
<221> primer_bind
<222> 157814..157832
<223> 4-104.pu complement

<220>
<221> primer_bind
<222> 172692..172709
<223> 99-12818.pu

<220>
<221> primer_bind
<222> 173072..173091
<223> 99-12818.rp complement

<220>
<221> primer_bind
<222> 180248..180268
<223> 99-24807.rp

<220>
<221> primer_bind
<222> 180874..180892
<223> 99-24807.pu complement

<220>
<221> primer_bind
<222> 184662..184680
<223> 99-12827.pu

<220>
<221> primer_bind
<222> 185138..185156
<223> 99-12827.rp complement

<220>
<221> primer_bind
<222> 190178..190196
<223> 99-12831.pu

<220>
<221> primer_bind
<222> 190643..190663
<223> 99-12831.rp complement

<220>
<221> primer_bind
<222> 191011..191030
<223> 99-12832.pu

<220>
<221> primer_bind
<222> 191441..191460
<223> 99-12832.rp complement

<220>
<221> primer_bind
<222> 195099..195116
<223> 99-12836.pu

<220>
<221> primer_bind
<222> 195568..195587
<223> 99-12836.rp complement

<220>
<221> primer_bind
<222> 203585..203602
<223> 99-12844.pu

<220>
<221> primer_bind
<222> 204095..204115
<223> 99-12844.rp complement

<220>
<221> primer_bind
<222> 210079..210096
<223> 4-24.pu

<220>
<221> primer_bind

<222> 210476..210495
<223> 4-24.rp complement

<220>
<221> primer_bind
<222> 210979..210996
<223> 4-27.pu

<220>
<221> primer_bind
<222> 211382..211401
<223> 4-27.rp complement

<220>
<221> primer_bind
<222> 215852..215870
<223> 5-400.pu

<220>
<221> primer_bind
<222> 216213..216231
<223> 99-12852.pu

<220>
<221> primer_bind
<222> 216253..216271
<223> 5-400.rp complement

<220>
<221> primer_bind
<222> 216708..216728
<223> 99-12852.rp complement

<220>
<221> primer_bind
<222> 221530..221549
<223> 4-37.rp

<220>
<221> primer_bind
<222> 221956..221973
<223> 4-37.pu complement

<220>
<221> primer_bind
<222> 225554..225572
<223> 5-270.pu

<220>
<221> primer_bind
<222> 225827..225845
<223> 5-270.rp complement

<220>
<221> primer_bind
<222> 229341..229359
<223> 99-12860.pu

<220>
<221> primer_bind
<222> 229770..229790
<223> 99-12860.rp complement

<220>
<221> primer_bind
<222> 237412..237429
<223> 5-402.pu

<220>
<221> primer_bind
<222> 237747..237766
<223> 5-402.rp complement

<220>
<221> primer_bind
<222> 1980..1998
<223> 5-390-177.mis

<220>
<221> primer_bind
<222> 2000..2018
<223> 5-390-177.mis complement

<220>
<221> primer_bind
<222> 4582..4600
<223> 5-391-43.mis

<220>
<221> primer_bind
<222> 4602..4620
<223> 5-391-43.mis complement

<220>
<221> primer_bind
<222> 10209..10227
<223> 5-392-222.mis

<220>
<221> primer_bind
<222> 10229..10247
<223> 5-392-222.mis complement

<220>
<221> primer_bind
<222> 10267..10285
<223> 5-392-280.mis

<220>
<221> primer_bind
<222> 10287..10305
<223> 5-392-280.mis complement

<220>
<221> primer_bind
<222> 39925..39943
<223> 4-58-318.mis

<220>
<221> primer_bind
<222> 39945..39963
<223> 4-58-318.mis complement

<220>
<221> primer_bind
<222> 39954..39972

<223> 4-58-289.mis

<220>
<221> primer_bind
<222> 39974..39992
<223> 4-58-289.mis complement

<220>
<221> primer_bind
<222> 41366..41384
<223> 4-54-199.mis

<220>
<221> primer_bind
<222> 41385..41403
<223> 4-54-180.mis

<220>
<221> primer_bind
<222> 41386..41404
<223> 4-54-199.mis complement

<220>
<221> primer_bind
<222> 41405..41423
<223> 4-54-180.mis complement

<220>
<221> primer_bind
<222> 42213..42231
<223> 4-51-312.mis

<220>
<221> primer_bind
<222> 42233..42251
<223> 4-51-312.mis complement

<220>
<221> primer_bind
<222> 67456..67474
<223> 99-86-266.mis

<220>
<221> primer_bind
<222> 67476..67494
<223> 99-86-266.mis complement

<220>
<221> primer_bind
<222> 69502..69520
<223> 4-88-107.mis

<220>
<221> primer_bind
<222> 69522..69540
<223> 4-88-107.mis complement

<220>
<221> primer_bind
<222> 72819..72837
<223> 5-397-141.mis

<220>

<221> primer_bind
<222> 72839..72857
<223> 5-397-141.mis complement

<220>
<221> primer_bind
<222> 76041..76059
<223> 5-398-203.mis

<220>
<221> primer_bind
<222> 76061..76079
<223> 5-398-203.mis complement

<220>
<221> primer_bind
<222> 81234..81252
<223> 99-12738-248.mis

<220>
<221> primer_bind
<222> 81254..81272
<223> 99-12738-248.mis complement

<220>
<221> primer_bind
<222> 83902..83920
<223> 99-109-358.mis

<220>
<221> primer_bind
<222> 83922..83940
<223> 99-109-358.mis complement

<220>
<221> primer_bind
<222> 91898..91916
<223> 99-12749-175.mis

<220>
<221> primer_bind
<222> 91918..91936
<223> 99-12749-175.mis complement

<220>
<221> primer_bind
<222> 95330..95348
<223> 4-21-154.mis

<220>
<221> primer_bind
<222> 95350..95368
<223> 4-21-154.mis complement

<220>
<221> primer_bind
<222> 95492..95510
<223> 4-21-317.mis

<220>
<221> primer_bind
<222> 95512..95530
<223> 4-21-317.mis complement

<220>
<221> primer_bind
<222> 96171..96189
<223> 4-23-326.mis

<220>
<221> primer_bind
<222> 96191..96209
<223> 4-23-326.mis complement

<220>
<221> primer_bind
<222> 97275..97293
<223> 99-12753-34.mis

<220>
<221> primer_bind
<222> 97295..97313
<223> 99-12753-34.mis complement

<220>
<221> primer_bind
<222> 98005..98023
<223> 5-364-252.mis

<220>
<221> primer_bind
<222> 98025..98043
<223> 5-364-252.mis complement

<220>
<221> primer_bind
<222> 98895..98913
<223> 99-12755-280.mis

<220>
<221> primer_bind
<222> 98915..98933
<223> 99-12755-280.mis complement

<220>
<221> primer_bind
<222> 98944..98962
<223> 99-12755-329.mis

<220>
<221> primer_bind
<222> 98964..98982
<223> 99-12755-329.mis complement

<220>
<221> primer_bind
<222> 103574..103592
<223> 4-87-212.mis

<220>
<221> primer_bind
<222> 103594..103612
<223> 4-87-212.mis complement

<220>
<221> primer_bind

<222> 104379..104397
<223> 99-12757-318.mis

<220>
<221> primer_bind
<222> 104399..104417
<223> 99-12757-318.mis complement

<220>
<221> primer_bind
<222> 106354..106372
<223> 99-12758-102.mis

<220>
<221> primer_bind
<222> 106374..106392
<223> 99-12758-102.mis complement

<220>
<221> primer_bind
<222> 106388..106406
<223> 99-12758-136.mis

<220>
<221> primer_bind
<222> 106408..106426
<223> 99-12758-136.mis complement

<220>
<221> primer_bind
<222> 108296..108314
<223> 4-105-98.mis

<220>
<221> primer_bind
<222> 108308..108326
<223> 4-105-86.mis

<220>
<221> primer_bind
<222> 108316..108334
<223> 4-105-98.mis complement

<220>
<221> primer_bind
<222> 108328..108346
<223> 4-105-86.mis complement

<220>
<221> primer_bind
<222> 108453..108471
<223> 4-45-49.mis

<220>
<221> primer_bind
<222> 108473..108491
<223> 4-45-49.mis complement

<220>
<221> primer_bind
<222> 109177..109195
<223> 4-44-277.mis

<220>
<221> primer_bind
<222> 109197..109215
<223> 4-44-277.mis complement

<220>
<221> primer_bind
<222> 114585..114603
<223> 4-86-60.mis

<220>
<221> primer_bind
<222> 114605..114623
<223> 4-86-60.mis complement

<220>
<221> primer_bind
<222> 115697..115715
<223> 4-84-334.mis

<220>
<221> primer_bind
<222> 115717..115735
<223> 4-84-334.mis complement

<220>
<221> primer_bind
<222> 122064..122082
<223> 99-78-321.mis

<220>
<221> primer_bind
<222> 122084..122102
<223> 99-78-321.mis complement

<220>
<221> primer_bind
<222> 123105..123123
<223> 99-12767-36.mis

<220>
<221> primer_bind
<222> 123125..123143
<223> 99-12767-36.mis complement

<220>
<221> primer_bind
<222> 123212..123230
<223> 99-12767-143.mis

<220>
<221> primer_bind
<222> 123232..123250
<223> 99-12767-143.mis complement

<220>
<221> primer_bind
<222> 123258..123276
<223> 99-12767-189.mis

<220>
<221> primer_bind
<222> 123278..123296

<223> 99-12767-189.mis complement

<220>
<221> primer_bind
<222> 123449..123467
<223> 99-12767-380.mis

<220>
<221> primer_bind
<222> 123469..123487
<223> 99-12767-380.mis complement

<220>
<221> primer_bind
<222> 126719..126737
<223> 4-80-328.mis

<220>
<221> primer_bind
<222> 126739..126757
<223> 4-80-328.mis complement

<220>
<221> primer_bind
<222> 128191..128209
<223> 4-36-384.mis

<220>
<221> primer_bind
<222> 128211..128229
<223> 4-36-384.mis complement

<220>
<221> primer_bind
<222> 128311..128329
<223> 4-36-264.mis

<220>
<221> primer_bind
<222> 128314..128332
<223> 4-36-261.mis

<220>
<221> primer_bind
<222> 128331..128349
<223> 4-36-264.mis complement

<220>
<221> primer_bind
<222> 128334..128352
<223> 4-36-261.mis complement

<220>
<221> primer_bind
<222> 128575..128593
<223> 4-35-333.mis

<220>
<221> primer_bind
<222> 128595..128613
<223> 4-35-333.mis complement

<220>

<221> primer_bind
<222> 128668..128686
<223> 4-35-240.mis

<220>
<221> primer_bind
<222> 128688..128706
<223> 4-35-240.mis complement

<220>
<221> primer_bind
<222> 128735..128753
<223> 4-35-173.mis

<220>
<221> primer_bind
<222> 128755..128773
<223> 4-35-173.mis complement

<220>
<221> primer_bind
<222> 128775..128793
<223> 4-35-133.mis

<220>
<221> primer_bind
<222> 128795..128813
<223> 4-35-133.mis complement

<220>
<221> primer_bind
<222> 130786..130804
<223> 99-12771-59.mis

<220>
<221> primer_bind
<222> 130806..130824
<223> 99-12771-59.mis complement

<220>
<221> primer_bind
<222> 133187..133205
<223> 99-12774-334.mis

<220>
<221> primer_bind
<222> 133207..133225
<223> 99-12774-334.mis complement

<220>
<221> primer_bind
<222> 135367..135385
<223> 99-12776-358.mis

<220>
<221> primer_bind
<222> 135387..135405
<223> 99-12776-358.mis complement

<220>
<221> primer_bind
<222> 139370..139388
<223> 99-12781-113.mis

<220>
<221> primer_bind
<222> 139390..139408
<223> 99-12781-113.mis complement

<220>
<221> primer_bind
<222> 157516..157534
<223> 4-104-298.mis

<220>
<221> primer_bind
<222> 157536..157554
<223> 4-104-298.mis complement

<220>
<221> primer_bind
<222> 157560..157578
<223> 4-104-254.mis

<220>
<221> primer_bind
<222> 157564..157582
<223> 4-104-250.mis

<220>
<221> primer_bind
<222> 157580..157598
<223> 4-104-254.mis complement

<220>
<221> primer_bind
<222> 157584..157602
<223> 4-104-250.mis complement

<220>
<221> primer_bind
<222> 157600..157618
<223> 4-104-214.mis

<220>
<221> primer_bind
<222> 157620..157638
<223> 4-104-214.mis complement

<220>
<221> primer_bind
<222> 172961..172979
<223> 99-12818-289.mis

<220>
<221> primer_bind
<222> 172981..172999
<223> 99-12818-289.mis complement

<220>
<221> primer_bind
<222> 180603..180621
<223> 99-24807-271.mis

<220>
<221> primer_bind

<222> 180623..180641
<223> 99-24807-271.mis complement

<220>
<221> primer_bind
<222> 180790..180808
<223> 99-24807-84.mis

<220>
<221> primer_bind
<222> 180810..180828
<223> 99-24807-84.mis complement

<220>
<221> primer_bind
<222> 190315..190333
<223> 99-12831-157.mis

<220>
<221> primer_bind
<222> 190335..190353
<223> 99-12831-157.mis complement

<220>
<221> primer_bind
<222> 190399..190417
<223> 99-12831-241.mis

<220>
<221> primer_bind
<222> 190419..190437
<223> 99-12831-241.mis complement

<220>
<221> primer_bind
<222> 191378..191396
<223> 99-12832-387.mis

<220>
<221> primer_bind
<222> 191398..191416
<223> 99-12832-387.mis complement

<220>
<221> primer_bind
<222> 195109..195127
<223> 99-12836-30.mis

<220>
<221> primer_bind
<222> 195129..195147
<223> 99-12836-30.mis complement

<220>
<221> primer_bind
<222> 203827..203845
<223> 99-12844-262.mis

<220>
<221> primer_bind
<222> 203847..203865
<223> 99-12844-262.mis complement

<220>
<221> primer_bind
<222> 210132..210150
<223> 4-24-74.mis

<220>
<221> primer_bind
<222> 210152..210170
<223> 4-24-74.mis complement

<220>
<221> primer_bind
<222> 210302..210320
<223> 4-24-246.mis

<220>
<221> primer_bind
<222> 210322..210340
<223> 4-24-246.mis complement

<220>
<221> primer_bind
<222> 210370..210388
<223> 4-24-314.mis

<220>
<221> primer_bind
<222> 210390..210408
<223> 4-24-314.mis complement

<220>
<221> primer_bind
<222> 211149..211167
<223> 4-27-190.mis

<220>
<221> primer_bind
<222> 211169..211187
<223> 4-27-190.mis complement

<220>
<221> primer_bind
<222> 215977..215995
<223> 5-400-145.mis

<220>
<221> primer_bind
<222> 215981..215999
<223> 5-400-149.mis

<220>
<221> primer_bind
<222> 215997..216015
<223> 5-400-145.mis complement

<220>
<221> primer_bind
<222> 216001..216019
<223> 5-400-149.mis complement

<220>
<221> primer_bind
<222> 216007..216025

<223> 5-400-175.mis

<220>
<221> primer_bind
<222> 216027..216045
<223> 5-400-175.mis complement

<220>
<221> primer_bind
<222> 216063..216081
<223> 5-400-231.mis

<220>
<221> primer_bind
<222> 216083..216101
<223> 5-400-231.mis complement

<220>
<221> primer_bind
<222> 216199..216217
<223> 5-400-367.mis

<220>
<221> primer_bind
<222> 216219..216237
<223> 5-400-367.mis complement

<220>
<221> primer_bind
<222> 216303..216321
<223> 99-12852-110.mis

<220>
<221> primer_bind
<222> 216323..216341
<223> 99-12852-110.mis complement

<220>
<221> primer_bind
<222> 216518..216536
<223> 99-12852-325.mis

<220>
<221> primer_bind
<222> 216538..216556
<223> 99-12852-325.mis complement

<220>
<221> primer_bind
<222> 221630..221648
<223> 4-37-326.mis

<220>
<221> primer_bind
<222> 221650..221668
<223> 4-37-326.mis complement

<220>
<221> primer_bind
<222> 221848..221866
<223> 4-37-107.mis

<220>

<221> primer_bind
<222> 221868..221886
<223> 4-37-107.mis complement

<220>
<221> primer_bind
<222> 225626..225644
<223> 5-270-92.mis

<220>
<221> primer_bind
<222> 225646..225664
<223> 5-270-92.mis complement

<220>
<221> primer_bind
<222> 229368..229386
<223> 99-12860-47.mis

<220>
<221> primer_bind
<222> 229378..229396
<223> 99-12860-57.mis

<220>
<221> primer_bind
<222> 229388..229406
<223> 99-12860-47.mis complement

<220>
<221> primer_bind
<222> 229398..229416
<223> 99-12860-57.mis complement

<220>
<221> primer_bind
<222> 237536..237554
<223> 5-402-144.mis

<220>
<221> primer_bind
<222> 237556..237574
<223> 5-402-144.mis complement

<220>
<221> misc_binding
<222> 1987..2011
<223> 5-390-177.probe

<220>
<221> misc_binding
<222> 4589..4613
<223> 5-391-43.probe

<220>
<221> misc_binding
<222> 10216..10240
<223> 5-392-222.probe

<220>
<221> misc_binding
<222> 10274..10298
<223> 5-392-280.probe

<220>
<221> misc_binding
<222> 39932..39956
<223> 4-58-318.probe

<220>
<221> misc_binding
<222> 39961..39985
<223> 4-58-289.probe

<220>
<221> misc_binding
<222> 41373..41397
<223> 4-54-199.probe

<220>
<221> misc_binding
<222> 41392..41416
<223> 4-54-180.probe

<220>
<221> misc_binding
<222> 42220..42244
<223> 4-51-312.probe

<220>
<221> misc_binding
<222> 67463..67487
<223> 99-86-266.probe

<220>
<221> misc_binding
<222> 69509..69533
<223> 4-88-107.probe

<220>
<221> misc_binding
<222> 72826..72850
<223> 5-397-141.probe

<220>
<221> misc_binding
<222> 76048..76072
<223> 5-398-203.probe

<220>
<221> misc_binding
<222> 81241..81265
<223> 99-12738-248.probe

<220>
<221> misc_binding
<222> 83909..83933
<223> 99-109-358.probe

<220>
<221> misc_binding
<222> 91905..91929
<223> 99-12749-175.probe

<220>
<221> misc_binding

<222> 95337..95361
<223> 4-21-154.probe

<220>
<221> misc_binding
<222> 95499..95523
<223> 4-21-317.probe

<220>
<221> misc_binding
<222> 96178..96202
<223> 4-23-326.probe

<220>
<221> misc_binding
<222> 97282..97306
<223> 99-12753-34.probe

<220>
<221> misc_binding
<222> 98012..98036
<223> 5-364-252.probe

<220>
<221> misc_binding
<222> 98902..98926
<223> 99-12755-280.probe

<220>
<221> misc_binding
<222> 98951..98975
<223> 99-12755-329.probe

<220>
<221> misc_binding
<222> 103581..103605
<223> 4-87-212.probe

<220>
<221> misc_binding
<222> 104386..104410
<223> 99-12757-318.probe

<220>
<221> misc_binding
<222> 106361..106385
<223> 99-12758-102.probe

<220>
<221> misc_binding
<222> 106395..106419
<223> 99-12758-136.probe

<220>
<221> misc_binding
<222> 108303..108327
<223> 4-105-98.probe

<220>
<221> misc_binding
<222> 108315..108339
<223> 4-105-86.probe

<220>
<221> misc_binding
<222> 108460..108484
<223> 4-45-49.probe

<220>
<221> misc_binding
<222> 109184..109208
<223> 4-44-277.probe

<220>
<221> misc_binding
<222> 114592..114616
<223> 4-86-60.probe

<220>
<221> misc_binding
<222> 115704..115728
<223> 4-84-334.probe

<220>
<221> misc_binding
<222> 122071..122095
<223> 99-78-321.probe

<220>
<221> misc_binding
<222> 123112..123136
<223> 99-12767-36.probe

<220>
<221> misc_binding
<222> 123219..123243
<223> 99-12767-143.probe

<220>
<221> misc_binding
<222> 123265..123289
<223> 99-12767-189.probe

<220>
<221> misc_binding
<222> 123456..123480
<223> 99-12767-380.probe

<220>
<221> misc_binding
<222> 126726..126750
<223> 4-80-328.probe

<220>
<221> misc_binding
<222> 128198..128222
<223> 4-36-384.probe

<220>
<221> misc_binding
<222> 128318..128342
<223> 4-36-264.probe

<220>
<221> misc_binding
<222> 128321..128345

<223> 4-36-261.probe

<220>
<221> misc_binding
<222> 128582..128606
<223> 4-35-333.probe

<220>
<221> misc_binding
<222> 128675..128699
<223> 4-35-240.probe

<220>
<221> misc_binding
<222> 128742..128766
<223> 4-35-173.probe

<220>
<221> misc_binding
<222> 128782..128806
<223> 4-35-133.probe

<220>
<221> misc_binding
<222> 130793..130817
<223> 99-12771-59.probe

<220>
<221> misc_binding
<222> 133194..133218
<223> 99-12774-334.probe

<220>
<221> misc_binding
<222> 135374..135398
<223> 99-12776-358.probe

<220>
<221> misc_binding
<222> 139377..139401
<223> 99-12781-113.probe

<220>
<221> misc_binding
<222> 157523..157547
<223> 4-104-298.probe

<220>
<221> misc_binding
<222> 157567..157591
<223> 4-104-254.probe

<220>
<221> misc_binding
<222> 157571..157595
<223> 4-104-250.probe

<220>
<221> misc_binding
<222> 157607..157631
<223> 4-104-214.probe

<220>

<221> misc_binding
<222> 172968..172992
<223> 99-12818-289.probe

<220>
<221> misc_binding
<222> 180610..180634
<223> 99-24807-271.probe

<220>
<221> misc_binding
<222> 180797..180821
<223> 99-24807-84.probe

<220>
<221> misc_binding
<222> 190322..190346
<223> 99-12831-157.probe

<220>
<221> misc_binding
<222> 190406..190430
<223> 99-12831-241.probe

<220>
<221> misc_binding
<222> 191385..191409
<223> 99-12832-387.probe

<220>
<221> misc_binding
<222> 195116..195140
<223> 99-12836-30.probe

<220>
<221> misc_binding
<222> 203834..203858
<223> 99-12844-262.probe

<220>
<221> misc_binding
<222> 210139..210163
<223> 4-24-74.probe

<220>
<221> misc_binding
<222> 210309..210333
<223> 4-24-246.probe

<220>
<221> misc_binding
<222> 210377..210401
<223> 4-24-314.probe

<220>
<221> misc_binding
<222> 211156..211180
<223> 4-27-190.probe

<220>
<221> misc_binding
<222> 215984..216008
<223> 5-400-145.probe

tgtagcagca acatcactgc ctgtggaggt ggtggaggta gaatattagc aggagtaggt 34140
aatgatgttg aaagggaaga aggaaaacgg ggtgtggggg gttgttcttt aaaaggaatc 34200
acattcctga agtatgaagg cacttttgg tcttaaagtg gatttttgt ttattttcag 34260
atgatgatgt acctattctc ttatttgaat ctaatggttc attaatatat actcccacaa 34320
ttgaaattaa tagtagtcac cacagcgcaa tggagaagag attacaagag atgaaggaga 34380
aaagggaaaa tctttccccc acctgtaagt aattagtttg taaaatgaaa attatgcaaa 34440
tagccgattc aattatggtg gaaagcttct tttttctttg cctagatatt ttaatgtttc 34500
ctggtagtaa cacattttga cttatttcat ggctggcttt gttttccaga aaatcttatg 34560
catcattaag atttttgaag catatgttgg gtgtatagta ttcttcaagt ttaaaatcct 34620
atttgttgta gctcctttgt aatttctatt atctttggaa ttttttcttt ctttttttt 34680
aaaaaaaaaa tgaatcatgt ctttttttt ttttctgaga tggagttttg catttgtcac 34740
ccaggctgga gtgcagtggc gcgatctggg ctcactgcaa cctccctagt tcaagtgatt 34800
ctactgcctc agcctcccga gtagctggga ttacaggcgc ctgtcaccac tcctggctaa 34860
ttttttttg ttttttgta tttttagtag agacggggtt tcaccatgtt ggtcaggctg 34920
gtcttaaact cttaacctca ggtgatacac ccgcctcggc ctcccaaacg gctgggactg 34980
taatccaggc gtgagccacc gctcctggcc gtgaatcatg tcttttgaag gaatttgctt 35040
tagattaatg tatctaagga atcagtttgt ttttcattat ttctttttatc tttaaaattt 35100
ttaattactg aagtgtaatt cacattttaa taaaacattt atcaaagtag ctaatagtaa 35160
aagttcatct tgatacccat ctaattgtac tcttctacct gggggtaacc tgtattttaa 35220
gtttaagtgt tttcccagat ctgtttcagt gtatcagata tctgtgtata catgaaaaag 35280
atacgggttt ggtttctgtg tggaggtgta atttctgttt tacctaaatt agataatgac 35340
atatgtatta ttatccgctt tatttactta agagtatcct ggagggtttg tttgcagctt 35400
agttgttgta gacctatttt tgttttaaga tgctcaaagt agtctacagt tttgatattg 35460
aaaatctatt ggtgggtatt tttttcccag ttattagaaa ttgtgttgca gtttttattc 35520
ttttttttaac catatggttt ggttgttctt gttttttttgt taagccattt tcctttctct 35580
agacataagt ctttccagct tcccaccccg acttttttact gttataaccc ctgcatgtgc 35640
ctacgtgaat ccttgtattt ctgagtactt cgtgtatttc aataatacta attcatacat 35700
gcagaatttg atttttttaaa gacatagagt ctccctgtgt tgcgcaggca ggacatgcac 35760
tcctgggctc aagtacttct gcctcaccct ctcaagtagc taggaataca ggtgtgtgcc 35820
acgatccctg gcttattgat agatatagtc aaattatcct tcaaaaaatt tgagtcatct 35880
tattgtcacc agttgtttat aagaatgccc ctttctccat acttggaaaa ctgaatggca 35940
ttagcctgta gccttttttca gtcggaagct tgaaaaactg gatctgttct tgaagttact 36000
tttgattaga agcaggttta agtgcctttt catattactg actgacttac cgaatgcagc 36060
ttttaatgtg atcaactatt acctcgctta attttatgtc ctttgtccat ctgtatcagt 36120
taaggttagt ttcggctgca tataacaaag acaaaaacca atgtgttaca atcgatagaa 36180
ttgcctttct ctgtcttgcc tagttcagaa gtaggcagcc agggctggga tgccattcca 36240
tggtgtcttt aagaaactag gttcccatct ttctgttgta cctgcctggc ttttcttgca 36300
aaatgtgtgt gcctcccagc taagccatct ccttttgaca gccttaccag acgtctatcc 36360
aatattcctg tctaattcca ttggctggaa tgtggtcata tggccacccc ttttgcaagc 36420
aagactgaaa tgtagtcttg actgggatgc attgctgtcc tgataaaatc aaagttctgt 36480
tgttaagaag aagtgagaat ggacattgag gtagataact agctgtgtcc caggtggaca 36540
tccaaattgt ttcagtgtgc aattatgtgt ataaactaat ttgccttaaa ctttacttttt 36600
tctattactt ggcagtgtta attctgctac tttactgcgt ccagtacagt ttaaaactta 36660
actgaaaatt ttatgtgtgc ttcccttcct tatcttggtt tattctcttt tttttgctga 36720
agttttctca gaaaagtatc cttttgagtc tctaaaaaat atctttggat ataagatcca 36780
aacatttctt ttgtttcttg actattgtat gaaccgcctt tgaagataat acttacgatc 36840
ttatttgtta agtcattgac atcctaagtg ttttctatga aacctctagg atttctcaac 36900
ccagcacagc tgacatttgg gtctgggtaa ttctttgttg ggggcactgc cctgtgtgtg 36960
gtaggaagct cagcagcatc cctgcctctc cccactaaca ctagcagtgt acctactgct 37020
ctccctcact ggcgatatcc aaaaatgtgt ccagacatta ccaaatatct gctgggaccc 37080
caacgtcacc tctggttggg aagcagtgct ctagttttag aggtaactat gatgagcatc 37140
cttgaagaaa aatccatgat tatcaaataa gaagactaga acagactgga aatgttcact 37200
taattctgtt gagcttctga ttagattcag gcaagttgac tttaagatcc cttctaactt 37260
tgtgattata ggatttaata gaatcaccta tgattaatag gaggacttcc tgctggcttc 37320
gtctgctaag aaatactgaa actttatcta atgcagtgtc ttggtcctgt ttttagcttc 37380
ccaaatgatt cagcagtctc atgataatcc aagtaactct ctgtgtgaag caccttttgaa 37440
catttcacgt gatactttgt gttcaggtaa aatttttatt ttcctttctg tgatatgttt 37500
aagttttgag aataatatga ttttctgatt tagaatttca tgtagcaact tctgatgagt 37560
aaaataatta gttaaaacta gaacttctaa atttcccccc gaaattaggt attataataa 37620
aattaaggca tgagttaaac ttccttttg gttcctatag gttttttttt cctaggcatt 37680
tgctttcttg ctacagaatc cattgctcta tttaaaaaat tattgtgaac gtatatgaac 37740
taatctgtat gcagtttaaa ctacatagaa ctgaggtcag agctaaggaa atgttgtttc 37800
acacaatgta taattaacac aaggaacctg ttattgaacg gggtcagtga agtatgtaaa 37860

<220>
<221> misc_binding
<222> 215988..216012
<223> 5-400-149.probe

<220>
<221> misc_binding
<222> 216014..216038
<223> 5-400-175.probe

<220>
<221> misc_binding
<222> 216070..216094
<223> 5-400-231.probe

<220>
<221> misc_binding
<222> 216206..216230
<223> 5-400-367.probe

<220>
<221> misc_binding
<222> 216310..216334
<223> 99-12852-110.probe

<220>
<221> misc_binding
<222> 216525..216549
<223> 99-12852-325.probe

<220>
<221> misc_binding
<222> 221637..221661
<223> 4-37-326.probe

<220>
<221> misc_binding
<222> 221855..221879
<223> 4-37-107.probe

<220>
<221> misc_binding
<222> 225633..225657
<223> 5-270-92.probe

<220>
<221> misc_binding
<222> 229375..229399
<223> 99-12860-47.probe

<220>
<221> misc_binding
<222> 229385..229409
<223> 99-12860-57.probe

<220>
<221> misc_binding
<222> 237543..237567
<223> 5-402-144.probe

<400> 1
tctccccaaa ttcatctgta gagtcaacac aatctcaatc aaaatcccag cagtattttt 60

ttgtgcaaaa tgagaagtcg actctaagat ttaaaatgaa atctgaagaa tctagaagat 120
acaaaataac cttgaaaaat aaagttgtag gacataaact atctgatttc atcacttatt 180
tataagctac aataatcaaa acagcatggt gctggcagca aaaagacaaa tagctcaatg 240
gaacacaata ggaagcctaa aatgaaacac atacatatgc aacacagatt ttgatgtaag 300
cacaaaggaa atgcagtaga gacaaaaata acttttttaat aaatgatgct ggaacatttg 360
gatatgtata catgcaaaaa aatgaacttt ggtccctatc ccataccgta tacaaaaatt 420
aattaaaagc agatcttatc ctttgagtcc agtaggttga ggctgcagtg agctgtgatt 480
acaccactgc attccagcct gggcaacgga gtgagaacct gcctggagaa aaaaaaaaaa 540
aagtagaacc tagacctgat atacaaccta aagcagtaat atttctagaa gaaatcctag 600
gagaaaatat ttgtgatcgt ggagatgaag aatctatcaa atactaaact ttttttacca 660
ccttgaccaa aagtaattgg tttatatact tcatcatatc atttaattct aaatctacag 720
agatcaatgt cactttctca gtaaaagtac gtgagtcttc aatgatgccc tgaactcaca 780
ctcccaagta aaccataaca ccatatttcc agagtagagt ttattagaac aataactggt 840
gataatgata aatattgatc aaagactgag cctaggaagt gggtttttg aggctgcata 900
tactcaaggc aattcttcag aaccacagag ggctcattgg atcctattaa aagctgagag 960
ttaatgaata aacagataaa acagagacct gagtagacgg tagtcgatat tcttgtacat 1020
gtattctacc tctagattcc atagaaagaa ctaaaagtac atgaatttca ctaccaacat 1080
ctccatcagt taccagctgt atcaccttgg atcagtcagg taacctcccg cgaatttgct 1140
tccggggcag gggatcgcgc tgcaggtttg agcctgggag ccggcagggt ggagcagttg 1200
gagggccaag cctttgagct ccaggggggg tggccgggac agtgggtagt gccagccgat 1260
cggcgtcctg gggattgcct gaatgtgagg tctgggttca ccccgcggtg acctgagtcc 1320
tgggatgccc ctacagtgat ttgctgcctc agggatccga agtctctttc attcccttac 1380
tggggatttg aggtctggag gtactcctgc gggggtctga gatctcgggg tcaccctgtg 1440
ggggtctgaa gcctcgggtc cccgctgggg tctgaggtat cagagtcccc tccgttgggt 1500
ctgaggtctc ggggtccccc atccccggga tcggaggtcc ggctccccgg agcaggcagg 1560
gcggtgcgtc tggccctgac agtaacgtgg cgcgccagcc ccaggtggtg tcgggctagg 1620
ggggcataac ggtgccgaaa gtccgcacaa agccgtccgc tggggtcccg ccgcgcccgc 1680
gaggcaatga ctgtgccccc tccccttcct gatcctcagc tcaggtgagc ccagatgagg 1740
cgccgggtag cttctaagtc actaatggaa atagaaggct aattcagggg ttaggggccg 1800
tcgtccttct tactcgcagg agaagagaaa aacccacggc ccagcagcca gaggcgcggc 1860
gaggcggaat cgggccccct ccccgggggc tcagctccct ccagcctccc gcctcaccta 1920
cagagaaatc ccggaaacgc ggattcagcg gagcgcggtg acggcggcgc gctcaccccg 1980
cgcatgccca gtgcccgcsc gcgccgccag gctcgcaagc accgcgtagg ccagctggcc 2040
ggatcccgcc gtctgtcatg gcggccccca tcctgaaagg tgaggtactt cctgctgcct 2100
gctccagcag cgggagtttg aggaccggca cccctcgtcg cgggcgcact cgggggatcc 2160
cgtgggagga gccccgctcg cccctccctc gctgcctgtc tcccccagac cccctgccgc 2220
ctccttcctc ccccgctgcc tgtcccccca aaacccccgg ctgcctgctt cgtctcccgt 2280
gctccctgtc cccccaaacc cccgactgcc tgcttcctcc cccgtactgc ttgtgcccca 2340
acccccgtgc tgctagttcc cctcaatccc ccgctgcctg ctccctcccc catgctgcct 2400
gtcccccaaa tcccgccttt ccccctacct gctttcaccc ctgctgcctt agtccctgga 2460
tctggggctc actggcaggc agagtcctgc cctccggaag ttggtgtggg gccctcctgg 2520
gtctggtcct gttcgacccc ctctgaggcc cacctggagg agcggcagtt gagtttctat 2580
gctaattgtt ccaataatag gagccgcctt ttactgcgga gtctttgtgt gccaggcgct 2640
gtgcttaggc tagtatggta ttgtctgatt tttttaaccg ctctatcaac tctcttatat 2700
cattgtacag gcagaaacta aggcattgga cgtttaggtg actctccctg tgtgtggcta 2760
gtcagtgctg acagggcctt agaccggagc tgctgtccta accagtatat gataccgcac 2820
gcagtcccac cctctgtgca cctggaagag cccaggagag gggaatagcg gacacgtgtc 2880
ttgtagagtt tgaccgtgag aaaaaagggg cctgtattgt ggggcctgca gtcataaaac 2940
ctcatagcca aaagtaaaga ctagaggctt tatacaaagt ctgtaatcag atgtggctat 3000
ttttctaatg ttagtatttt gttaaattaa cctggttttc ttttagcgtt acccccaatc 3060
attgaccaac ggcacacctg gaaaatgctt ttaaacatca ggttttgaga agaggatatc 3120
cactagaaca ggggtccact cactatgccc cccaggccat atctagcctg ctgcctgttt 3180
ttgtaagggt ctacgagcta agaatgtctt ttacattttt aagtgatttt aaaaaaaggt 3240
caaatgaaaa attatatcac attcacattt ccttctccat aaataaagtt ttattggaac 3300
acaggccggc ccgttaatat attacctatg gttatgtttg tgccacaacc gtgaagttga 3360
gtagttgtgg caaatactgt attggccaca aagcctgaaa tatttaccat ctgtctcttt 3420
acagaaaata ggtttctgca ctggaaaaat taagcgtaag aatttgggga aagcaactaa 3480
ttttacaaat gtaaactctc atgtattgta tgggtacagt tgttctttgc ttaaaatttt 3540
aataaattcc actgaagcta ttttgaaaag gctttcagta gaaatttatt tatgagacag 3600
agtcttactc tcttgcccag gctggagcgc agtgatgtga tcacataata gctcaagcaa 3660
ttctgcttca gcctcctgag taacttggga ctacaggcac taccatgccc ggttatttttt 3720
atttttattt tttagtttat tatttttttg tagagccagg gtctcactat gttgcctagg 3780
ctggtcttga attcctagcc tcaagcaatc ctcccgcctc caccttgcaa aatgctggga 3840

ttacaggcat gagctacttt gttcagccag tagaagaaac ttcatttact tttcttattt 3900
ttgaggcaag gtctttctct gctgcccagg ctggagtgca atggtgcgat cataactcag 3960
cttctacctc ctgggctcta gggattctcc cacctcagct tctccaccct acccaccccc 4020
atttcccacc cagtagctgg gactacagcc actcgccacc attcctggct aattaaaaac 4080
aaaatttttt ttagagacag ggtttcacta tgttgcccag gctggtctca aactcctgtg 4140
cccaagtgat cccactgcct tggccttcca gagtgctgca attacagcat gagccaccac 4200
acctggccag tagagtaaat ttttgtttta ctttttttctt ttttttattt ttgaaacggg 4260
tctcgccctg tcacccaggc tggagtgcaa tggcgcaatc tcggctcact gcaacctctg 4320
cctcccgggt tcaagtgatt ctcctgcctc agcctcccag tagctgggat tacaggtgcc 4380
cgccaccatg ctcggctaat tttttgtatc ttttagtaga gatggttttt caccatgttg 4440
gcccggctgg tctcaaaccc ctgacttcgt ggatccaccc acttccgcct cccacagtgc 4500
tgggattaca ggcgtgagcc actgtgccgg cctcggttta ctcttaaatg taaatagaac 4560
aaaatctatt gggcagggga tgctggaatt tcaaatgtat rtttcatgtt catatcttgt 4620
tttcagatgt agtggcctat gttgaagtgt ggtcatccaa tggaacagaa aattattcaa 4680
agacatttac aacacagctt gtggatatgg gggcaaaggt aagacactta ttttgctgtt 4740
gattcatatg acagtcttct gattggtaaa aagttacatt tgcattttct tattttggga 4800
gtttttactt agaatctgga cgaagcaatg ggtaagcggt gggagaaaaa agagccaaag 4860
tgtgaagaat ttagaacagt aggactttca gaactcaatg cctgtgggca ttgagtgagg 4920
aggaggaacc taggatgaaa tgctggattc ttacactggt tacttgaatg catagtgcta 4980
ttaagcaaag tgaggaatac aggaaaagga acaggtttct aagggaaaaa ttgtaaattt 5040
gggcatactg aaaatatctg ttagatattt ggatatacaa gtctggagct tggagtgttc 5100
aaggctagag atgatgatct agggggtcag gaccataggg gtcatgtgaa gtcacaggtg 5160
tggacatcgt cccatgtcag gcatggttag gatgaagagt ggtgacagag gagcgttgtt 5220
cagtattcaa ggacaggcga tgggagcagg gacccagtga cagagggaga gaagaatgcc 5280
aggagaagga gaaaggaagt gtggaagtca aagtagggag taatttttt tttttgagac 5340
ggagtttcgc tctgtcgcta ggctggagtg cagtgacgcg atctcagctc actgcaatct 5400
ctgccttctg ggttcaagcg attgtcctgc ctcagccttc caagtatctg ggactacagg 5460
cacatgccac catgcctagc taatttttt ttttgtattt ttagtaaaga cggggtttca 5520
ccatgttggc caggatggtc tcaatctcct gatctcgtga tccgcccacc tcggcctccc 5580
aaagtgctgg gattacaggc atgagccacc gagcccggcc aggagtaatt ttttaattgc 5640
ctttcagaac tagaatggag taatttttaaa gatagaattt ttaaaaacta cagaaagttc 5700
aagaaaaata ggatgggcaa atgtactttg gatttgaaca ctgtaaggtc attgctgaac 5760
ttagtgcagt tttcagtgaa atgggcagga atcattgagc tatgaggaaa tggagatagc 5820
aaacaatttg ccttattcaa ggtttcttag tatagccatc tctgttatca gatttactat 5880
cacgtactgc ttgtgttcag gtagcctcta tttgacttaa taatgtcctt gataccaaat 5940
aggtatcttt tgcccacgca cactaaaccg atcactttga tgacgggttt tacaaaaggg 6000
aaaagattca ttcacaggga agcccagcta ggaggcagaa gagtactcac atcttcattc 6060
ccaaagataa ggcttaggga tatttatcag ttagggaagt agggtgatct aagctgtggg 6120
gaaaaatgaa gtacatgatc tgcacaagca tagttgggat tcatggaatg catgtttaga 6180
aaacaggcat tattaggagg ccaaggcagg cggatcacct gaggtcagga gttcgagacc 6240
agcctggcca acatagtgaa accccatctc tactaaaaat acaaaaaaaa gccaggtgtg 6300
gtggcacaca cctgtagtct cagtgattcg ggaggctgag gcaggagaat cgtttgaacc 6360
tgggaggcgg aggttgcatt gagccgagat tgcaccactg cactccagcc tgggcgacgg 6420
agtaagattc tgtctccaaa aaccaaaaaa ataggcacta gtaggatccg atggtgaaga 6480
ttttggcctg atgtcaaaag gtcatttctt gggcatttac acaggcctgg ttgaagagtt 6540
ggtggttgca gcctgtttga actgtacggg tgctgcccca agttcctgaa aagtaactta 6600
agcaactgtt accgtggtga catatccacc agaagttttt atcttataag gaagccagtg 6660
aaggttatag catttagtag tatgacttgc agctatatag aaataaataa ataaataaca 6720
aaaagcaagt gaccaaaagc aagcaaggca ggttaaattt ggcagaacta attttcagcc 6780
gtaaagtgca agagtgatga tgctggcaat tcagatatgc cagagaagcc ttaaggtgct 6840
ttaagtgaaa aggtgaaagt tctccacttt aaggaaagga agaaaattgt gtgttgaagt 6900
tgctaagatc tacagtgaga acaaatcttc taatcttgaa attgtgaaga actatgctac 6960
tgttgcagtc acaccaaact gcaacagtta cagccacagt gcgtgatttt tattataata 7020
cattgctaca attaccctat tttgttatca ttattgttaa tctgtgccta atttgtaaat 7080
aaaacttcat tgtatatgta tgtataggaa aaaacagtat ataacctgtt cagtactagc 7140
tcaggattca ggcatccact gggaggggtt gggggcggga cgcgggcatg tcttagaact 7200
taacccccgt ggataagggg gaactaatgt gctcttatag ggagtttagt tatgaacaaa 7260
tcctgtttat gtccttgtct ggcatttggg aggggctgac tgataggctg agtgaaagag 7320
aaccattaaa aatgggagaa aagataatcg aaggcagggc taggttaggg tggagcaaga 7380
gagctgctgt gggtataaaa cttaagaggc gctcaccacc aggcaaagag tgggtgctac 7440
tgaataccct aagagccttg tttgacctcc ctaatgcctg tcttgagtaa gaggtcagtg 7500
gagaggaatc cgaatatagg agcagggcct gcactgcagg aggggagaca tgccccctgt 7560
aatacactgg aatgtaggaa cccgagggag tctgcatgtt gcacatgcct aacatttact 7620

tgggatgagg aggaactact gtgaataaga aaaaagccgt tagacaagtg agttgacaag 7680
gtggtttgag ggtagcatta agatcttaga tcttttagaa ctttttggtt tcaccttta 7740
tttcaaaaat tggcaaacag ttcaaagaat agtgaataca gatcaatagt cgttaacatc 7800
gtaagatttg gatatttgta gtactcgtaa tccgggatta tcttaagcca atttcaggat 7860
ttgagatgat ttaaaaccag actacaggcc tgtgagggtc attataactt ctgattcacc 7920
cttaatctag atgcagctct ttgggtctca gcgcaaggtg taggggtttt accaaacccc 7980
cttgtctctt tgaggcttct caattttcgt cctgtttagt gtgtactaaa tttgataaaa 8040
gccttgtggg aagatggtct caaatgctag actcatctct ctaggtgtca gtcttaatct 8100
agaatctcag cccggtaatt cttaattgcc ttgatagctc ccatggactt gatgggggtg 8160
tggaaattga gagagagaga gagttataaa agtaatacat atttattgtt taaaaagact 8220
aacgggcagt gccatgaaat tcacaatgaa aagaaagaga aaccagcaac gctttgcagt 8280
acatttcctt ttccattttt caaagacagc tactttcaaa tcatctgttt cttttggtat 8340
ttaccttcat atttccaagc attgtacata tattacttca gtataattga atgctataaa 8400
aatcatgcag atgagttctg cttctggaaa ggatacataa aaggtaaaat ttttgacacc 8460
atgactgtct gagcatgcct ttattatact gttacctttg actaatattt tagctgagtg 8520
taaaatgcta gcacaaatat tatttttcct tgaaggtatt tcccattgtt ttctcaattc 8580
cagactgctg ttgataagac tgattcagtt gtcacttatc attgtttgca tgtgatgtgt 8640
ctctatcctc ttcaccttga ttacccactc ttttaatatt tttccctttc gccaaccatg 8700
ctggaattct gtgataagct tggtgtagtg ctgttttcgt tcttgtgccg ggcctttgtg 8760
ggggattctt ttgatctaga atgcatatcc tttagtttga gaaacttttc tttgattatt 8820
tctttaataa tattttctcc attttgtgta ttcctgtatt ctttaacttc tgttgattgg 8880
ctgttggatc tcctggtctg agctcctgat gttcttgcct tttgtctcct gttgtccgtc 8940
ttctggttct tctcttctac taccagtgag cttttctcaa cttcattgtc tgatatttct 9000
gtagaaaatt tttttactta tatcatcttt tcttaatttc caagagctct ttaggatcct 9060
attaaaaaat aatcttctga tcatgttgca tgaatacagt atcttttgtt tttttttttt 9120
tggagatgga gtcttgctct gttacccagg ctggagtgca atggcacaat cttggctcac 9180
tgtaaccgcc acctcccggg ttgaagtgat tctcctgcct cagcctcccg agttgctgag 9240
actacaggca cgaacctcca cgcttggcta atttttgtat ttttagtaga gacagggttt 9300
ttccatgttg gccaggctag tcttgaattt ctgacctcat gatccacctg cctcggcctc 9360
ccaaagttct gggattacag gtgtgaacta ccacacccag tttcctttgg ttttaattag 9420
ctgaattttt ccaacttttt gaatgattgc acttattttc aaccttctta ctttgtattt 9480
atgcatttaa gattacaggc gtccgccacc ttgcacccgg ataatttttg tatttttagt 9540
agagacaggg tttcacgagg ttggctaggc tggtctcaaa ctgctgacct caggtgatcc 9600
acccgcctcg gcctcccaga gtgctgggat tacaggcgtg agccaccatg cccagccatg 9660
gatacagtat cttaagatat gaggtatttt taattttggt taaatatgtg ttctgttttc 9720
tctgttgcct ctgaatttca tttggtttta tttttttgat gtagaaagct tttctgaaat 9780
gtccattatt atctgactct ttccatcttt aaaaatgtgg tgccttctca tggccacatt 9840
ttctcttctg tcctctttat ccttgcaggg ctccaactct attctttcag taacacttca 9900
gagggttttt agagggagta gatgtgaact tgtgtgtatg attcaccgtt gtaactggaa 9960
cagatatgtt ttaagcagcg ttatacattc ctttgagtgt ttctctgtca gattttgaga 10020
aacagaattg ctggggtaga ggtttttga tcagttgtag ttaagttgtg aatgaacagt 10080
aatgtacatt ttgttttctg cattttgtct acaggtttca aaaactttta acaaacaagt 10140
aactcacgtt atcttcaaag atggctacca gagcacttgg gacaaagctc agaagagagg 10200
cgtaaagctc gtttcggtgc tctgggtkga aaagtaagca gtttctctct tactttttt 10260
ccttaagtat ctagtattga aaatgkgtgg agatattttt cacaggtcgg agaaccagat 10320
aaagtttgat tttcatcttt tctctgcctc ttacctcacc aagtaattta catcctccag 10380
cctcaatttc tgtggttcaa aaatggtcat gctataatac ctaactctgc ctaggggaa 10440
aaggagcctg caggtcctga agctgggtat gcaaggtgga cttaggaagc aagagggaat 10500
gtgatgaagc agattgtgtt agtcagcaag cgctgctgta acaaaggacc acagaatggg 10560
tcgcttgagc aacagaaaag gactttctca caactctgga ggcaggaagt ccagtatcaa 10620
gttgtccaca gggttggtat cttctttttt ttttttttga gacaaagtct tgctctgtca 10680
tccaagctag agtgcagtag ctggatcttg ggtcactgca gcctcagcct cctaggctca 10740
agtgattctt atgcctcagc ctcccaagta gctgggattc atctcaacct ttgcctcctg 10800
ggctcaagtg attctcctgc ttctgcctcc cgagtagctg ggattacagg cacgcaccac 10860
catgcctggc taatttttgc atttttggta gagacggggt ttcatcatgt tggccaggct 10920
ggtctcaaac ttctgacctc aggtgatcca cctgcctcgg cctcccaaag tgctaggatt 10980
acaggtgtga gccaccgtgc gcggcccaca cagttttgat tacagtaaat ttgtagtaag 11040
ttttgaaatt gggaagtacg agtcctgtaa cttgtttttc attttcaaga ttgtttggct 11100
attttgattt gagttccttg ctaagattgt ttggctgtct tgattgggtt ccttgcattt 11160
ctatatgaat tttatgatca gtgtgtcaat ttattcaaaa aacaaaaaag gcagctggga 11220
tttggtagga ttgtattgaa tctctaatta gggaagtgtt cataatattt aatcttccag 11280
tccatgaaaa tgggatgtgt ttctttttca ggtctcaaat ttccttcagt gatactttct 11340
agttttcagt gtacaagttt tttacccccc aggttaaatt tattcctaac ttttttgttc 11400

attttcatgt gaatgaaatt gttttcttaa ttttttaag ttgttagctg ttagtgtata 11460
gaaatgcagg tgattgttgt atgttgatct tataccctgc aaatttgctg aacttgttta 11520
ttagttctaa atatatttgt gggttcctta gcattttcta tatgcaatgt tgtgtaattt 11580
tgtaaataga gatagattta cttcttcatt tctagtctgg ccgcatgtta tgtcatgtca 11640
tgtcatgtca tgttatttgt tctggccaga acctccagca cagtgttgaa tagaagtggt 11700
gagaatggac gtccttgtgt tgttgctcat ctttggagaa aagctttcag tatttcatta 11760
tttcgtatga tggtaactgt ggtttgtgta aatgtccttt tttaggctga ggacgttccc 11820
attcccttct gttgcaggtg gtttgtttgt ttctgattat taaaggaagt tagatattgt 11880
tgtcagatgt ttttctgcat gactgatcat catgtgattt ttgtccttca ttatattaat 11940
gtggtgtaat tgaggggttt tgtgtgttga agcaaccttg cagtcctagg ataaatccta 12000
cttggtcatg ttgtatacgt tgtcatcttg ccttttttaat tgttggaaca ggccggatgc 12060
ggtggctcac acctgtaatt ccagcacttt gggaggccga ggggggtgga tcacctgaga 12120
ttaggagttt gagaccagcc tgatcaatat ggtaaaaccc tatctattaa aaatacaaaa 12180
attagccagt catgttggcg tgtgcctata gtcccagcta ctcgggagat tgagacagga 12240
gaatcacttg aacctgggag acggaggttg cagtgaacca agaccacgcc attgcacttc 12300
agcctgggtg acaagagcgg ggaaaaaaaa aaaagagtaa ggggtccttc tctggctttt 12360
gtcaagattt tctcattatc tttcactttc agaagtttaa gtgtctcgat atggttttt 12420
aaaatttatt ttgtttagat tttacagaac ttgttgaatg tgtagttgcc tgtgttttct 12480
acatatgatt cgtttttggc cattatttct tcatctatct tttctgcccc gtttcctcct 12540
catttttct gattagctgt atatcttttt ctaattagct gtataccagg gattggtaaa 12600
gttttctgta aagggacaga tagtcaatat tttaggcttt gcgggccata tggtctctgc 12660
tcaacagctc agctctcttg tggtgtgaaa ggcgtaatag aaaataagta aacaaatgct 12720
tgtgtctgtg tggcagcaaa cttataagtc tggcaggaag ccaggtagtt tcccaatcct 12780
ttctgtgtaa tacacctttt aaacgtttgc atttgtccat agtgctctcc ttcttcttca 12840
ttattgttca gtctttttttt ttctcctcag attgatcttt tttcaagttc attgactttt 12900
ttctccatca aatccattct gctacttagt acctttattt gagatatttt tgatttctaa 12960
catttctagt tggtttttgt atacattctt tttatttcgg tttcatgtga gatttctcac 13020
ctttggtttt cctgtcttcg gttatgagtg tattttctat tacctcgatg agtgtagtta 13080
taaatagttg tcttaatgac cttgtctgat aattttgagg ttggtatctg ttttgttttt 13140
gtttttcttt gatagtgtgt cacatttttc tggctcttca tatggcaaat aatttgaggt 13200
tgtattatgc acgttgtaaa tactatgtag actctggatt ctttttctatt gtcgcaaaga 13260
gcatgaggtt tttgttttag caagcagtta acttgtcgtt aaaatgaaac gcacactgtc 13320
attatgtggg cagttgctta gatgcgccct ttaagcctca ggtgcaggct gatttgtttg 13380
cctcaaacac atgttgttca ggggtcagcc agagacttga acttctatac tcagaatttg 13440
gggtttctcc tatggttctc ttacttcctg agtccttacc tcatttctct agtagcccta 13500
gctgcccagt ctccttcccc tggtctcttc agcgagaaag gaggccggag cttctgcttg 13560
agtgcttgct gcgccacacc agctccctca gagactgtgg ctgcctttag gggacagaca 13620
gaaaaagtgg tgatgcccag attcttttgc ttccttttaa aatttgcctg ttctttcttt 13680
ttcttttatt tccctccagc tttcaaagct ctcacatagt tggtttattt tattttattt 13740
ttcctgtatt tccaggacgt atagcttata gttcacctat atttgtatat tggtttgtta 13800
ggcataaaca gaaatggaac ttagtatgtt atttttgaag catctgatgc cagtctaatt 13860
cttcttccct tcaaaattat ttgatctttt tggagactcc ttagggatat tttttatttt 13920
atcatttttt ttttgagacg gagtctcgct ctgtcgccag gctggagtgc agtggcgcga 13980
tctgtgctca ctgcaacctc ctactccctg gttcagcgat tctcctgcct cagcctcccg 14040
agtagctggg atcacaggca cgtgccacca cgcccagcta atttttgtat ttttagtgga 14100
cacggggttt caccatgttg gccaggatga tcccgatctt ctgacctcgt gatctgcctg 14160
cctcagcctc ccaaagtact gggattgtag gcgtgagcca cagtgcccgg ccaggatttt 14220
ttttttaaga ctcatggctt tactgtaata tgttttgaat tgatcattcc agttctggct 14280
tggccttttc aacagattca ggtctatatt tctgcaaaag tttctgggaa ttatagtttt 14340
aaatattctg cttcgttgtt ttgcttttct tctgggactc caattatgtt tacgttgggc 14400
ctgcttagct atcttttatt tcagtcactt tgacttcaac ccttttatat ttatatacac 14460
atacacacac acacacacac acagacacac acacacacac acacacacac acacacaatt 14520
tttaccccaa atacttattt gacagtattt gtttttgttt tttgaagaca gggtcttgct 14580
ctgttgccga ggctggaatg caatgactca gttgcagctt actgcagcct tgacctctaa 14640
ggctcaatca gtcctctcac cccagccctc cctagtggct gggactgtag gcatgtgcca 14700
ccatgcccag ccattaaaaa gttttttttt ttctttttttc tttgagatgg agtcttgctc 14760
tgtggcctag tgcagtggcg caatctcggc tcactgtaag ctctgcctcc caggttcatg 14820
ccattctctt gcctcagcct cccgagtagc tgggactaca ggcgcccacc accacacctg 14880
gctaattttt ttttttttg tatttttgta gtagagatgg gattttaccg tgttagccag 14940
gatggtcttg atctcctgac cttgtgatcc acctgccttg gcctcccaaa gtgcaacccg 15000
gcattaaaga attttttta tagacatggg atcttactat gtaggccagg ctgggctcaa 15060
gtgatccact cactccagcc tctcaaagtg ctgggattac tggtgtgagc cactgcaccc 15120
agctgataat atttgattca agttcaaggg ttttgttata ttcttcagtt ttgtgtttgc 15180

ttttatttta gggagtgtga tgggttttcc tcagctgaaa tgatttgctt tttctttgtt 15240
tttttaaaat agatttttaa aatggatgta gtctattcta tttccattca ttgcataggc 15300
caggcttgtg gccagagcgt cctcttctgt cagttctgct gtcttgcata gtttctttta 15360
taggtgacgc tggtgaggga gggaggaggg aggggctcgt gtatctcgtt tgcgttttgt 15420
ttctatagga tccttaaatg tttttctctt agtttcttct ttttttcact gccattggtt 15480
caagggctgc cactcccccc agaactgatg tttttcagag cctgcctgtc ctagtcttgc 15540
tcccattcag acccccttccc tggagtgggt gctgtgagct gtgtgggttc tctgttgtgg 15600
cagttgtgct gggtgtcctc tttctgagac ttcttttacc tgtgcttcat gtaagttctc 15660
caggctgtac tacttttttat ggagtcttaa gcgtattctc cccgactttc tgcatccata 15720
gacttgcagc tgtgttggaa tttgattatt tttctactta taggtcatct gaatttgcgc 15780
tgttatctcc gtgtcagtga gaatgtaggt catatgtgtc ttttatttaa gtttcttttt 15840
tattttctgc tttttttttcg gggagggaat ggggtaagac tcagtatcag ccagccatca 15900
ttgttttctc tacctcatct tcttatggag tccattgaaa tggcttattg atttttatct 15960
caaaatcgat ctctcataga tctttatctc tgctgttaca gtcgagacaa gtatcatgtc 16020
ttgcttcagt tactgtagca gcctcatgcc tgtctgtttc attttgtttc ttatacataa 16080
gcaaatgtaa ccccttttgt taccagtgga aggtatccaa gttaccggca gcaaacacgt 16140
atgggtttgc agcaacttca gttcttgctt cctcaaaaga aagaattcca cggaggagca 16200
taaggcaaaa gaagagcctg acgcaagggt cagagcagga gcagaagttt atttaaaagg 16260
cgtcagaaca gaaagaaagg aaagtacact gggaagagtc ccaggcgggc atggaggtct 16320
aatttgatgt ttaaccttga tcctgggatt tgtaggctcg ccctttccg cagttcttcc 16380
cttagggtgg gctgcccgca tgcacagtgc gggaattgag cacaggcagc ttgtttagga 16440
agttgtgtgg gtgcccatct gaagctttct tcccgtttct ccgccatttt gtctcttaat 16500
gtgcatgccc gggaaatggc ctctccctgg cgtctgcatt cagttaacac tttagcacaa 16560
caggtgtgga ctgtcaggaa atggcctctc cctggctctg gctgccaatt tatcactttt 16620
agagaggcaa tgtgataatt gttgagctat cacccaacat tcctagtggg tggtagaggc 16680
ctctcctgcc gggcttatgc ctaactacct gtgatacttc aacacatgga tcagctttat 16740
ccttctgaca aaatggctta gagttcagtg gtctatagca gagaatggcc aactatcatc 16800
ccccagccaa atccaccctg ccatactgtt tatttttttt taatggccca tgaggtaaga 16860
atggttaaga gaaaaaaaaa attcaaatgt ttactatttc atgatattta cattatatga 16920
aattcaattt tagtatccat aaataccgtt ttattggaac acaggcatgt tcatctgacg 16980
atgtagtcag tggctgcctc tgtactacag ctgtagattt ggatcctgtg gcagagacct 17040
tacggcccat gaagcctaag gcattcacta ctttcccctt tacagaagtt tgctgaccca 17100
ggtccagtgt gctgcatgat ggtccccttc ccttccattg tcagctgctc ccctctccct 17160
tgtttgcgtc ttccaaatgc tctaggcttc cacgtcccca aggctacact ctttctgcct 17220
ttagttcttg gcctgtgctg agaactctgc cccgtcttcc tgattctaaa cccagttttg 17280
tagtcagctc ctttatacat gttgcattgc aaggtcgctt tatcagaaga gcttcctctg 17340
tccccagttc acagttcaag ccctatttgt tattctctgt ctcagctcct tttttcctgt 17400
gtgtactatt aaaacttatt ttgttcattt gactgcttta tctgtctgtg tatctaatca 17460
tgcattttgt ctttctattg taatgtggat tccaagagca gctacctgtc tgtcttattt 17520
atggttgtgt ttctagtaag tctaacattc atctggctca tagtagatgc tcagtaaata 17580
tttgttctaa caaattatga acaaaggaaa atttagttaa gtggcgtaga gatactagag 17640
aaaatatcat gggggaaaat gatttgaaaa aaactacatt ttaaaagtcg tatagaaatg 17700
tggaggggag agtgcagaaa cagagacctt tactagaagc ttgaagtaaa tggagatgca 17760
tggacaaaat taaaatagta gccatttctg tacctaatag ggcctctcag ctaaccctac 17820
agtggggatg gtcactggta gtgtgttctg ctgagagtta gggattctta ctctgctttg 17880
ctggcccagc ccctgactca ttctctatcc cctttctctc tctctctatt tctgcccacc 17940
actaacccca gcctttctca aggggctcat gcagacccca taatacttgt aacttcgtta 18000
tccaaaagca aagttttctt tttcttttct ggagactgag tctcactctc ttgcccaagc 18060
tggagtgcag tggtgcgatc tcggcttact gcaacctccg cctcctgggt tcatgccatt 18120
ctcctgcctc agcctcccga gtagctggga ctaccggagc ccgccaccac gcccggctaa 18180
ttttttgtgg ttttagtaga gacggggttt cactgtgtta gccaggatgg tctcgatctc 18240
ctgaccttgg gatccgccct cctcggtctc ccaaagtgct aggattacag gcgtgagcca 18300
ctgtgcccgg ccaattttta tatttttagg agagacaggg tttcaccatg ttggccaggc 18360
tggtttaact cctgacctca ggtgatccgc ccaccttggc ctcccaaagt gctaggatta 18420
caggtaagag ccaccgtgcc tggcaaaagc aaacttttaa ggtcctcaga agctcaaaag 18480
tgaacttaat cttttggcat ttttcttttc ttttttttt ttttttttt ttgaaactga 18540
gtctcgctct gtcgcccagg ctggagtgca gtggtgcaat cttggctcac tgcattctcc 18600
tgcctcagcc tcctgagtag ctgggactac aggcgcccgc caccacgcct ggctaatttt 18660
tttgtatttt tagtagagac ggggtttcac cgtgttagcc aggatggtct ccatctcctg 18720
atcttgtgat ccgcccgcct cggcctccca aagtgctggg attactggca tgagcccctg 18780
cgcccggccc atacacttta gtcaacttt tattacaggt catttttttg cctgtacatg 18840
cagatacatc ccactttata tatataagaa tattttgtag tagctgtcca gtaatttatg 18900
taagcagtgt cctattggtg attgaagttt ttcatttctt agttatttt ttcaattaga 18960

aatattacag cattgagctt ctgtatgtat tacctttttg cgggtgataa atctttccat 19020
aggtttaaat ccccaaagtg ggctgttcat ttctgagagt ttacacattt aaatatgata 19080
gatgctgcca aattatcttc tggaaggagt gtactggttt ccattctcac tggaattatc 19140
aaaaaaatgc atgtttccca ataccttttgc taatgttgtg agttatcagt tctttttttct 19200
aatttgtaga agaaaaataa tagtttttat ttgcatttct ctgacttttа gtgagcttga 19260
attttcttca gcagagcata gagataagag ccaaactgac ctgcattttt tatgtcacgt 19320
ctgtcctttc ttggtgaact gcctgccttt ccaatgcagt agctcatggt ttccactgaa 19380
aatgtgaaca ttaacttcat aaggtcacta ggtgtcacta gaatcccatt ctgttgggtt 19440
ccttctggga gtgttcattt taagatcaga tggcaattga taaaattctg acatttcctt 19500
tggatgtaga aatttttacc ttgaagaaag aatacataaa gttgaaataa aggtcagctt 19560
ggcccccact ctaagttctg ttgaagacaa tttatcattt ttaaacaact gcaaactaac 19620
agctaggtgg ggaatacggt tcacaggctt tgtccttgct aggctgagag ttggttgctg 19680
accgaagcca tcacccccctg catttagtgt ttgctggaaa cagggacata ttcctgcata 19740
accacaacac aggccgacat taggggctta ccacggctcc tttcctcccg gaatcctcag 19800
actccattcc tatcctacca gcagccagct ccacttcccg cctcctcagc cttctcaccc 19860
tgcagccatt ccttagtctt tcactggctt ttgtgacttt gacactgttt aaggtcactg 19920
accagtgata ggaacgtccc tcagtttgga acggtctgat gtgtcctcct aatatcacat 19980
caatgtgtaa cagtggatgt gtagccattt aggactgggc aaattactca actgctgggc 20040
tctaggttcc tccagtagct cctgagttaa cttcctacgg ttatttagtg ctagaccaca 20100
gaagttcgct ctctgctggc agagcactgt tgtgcagact tctctgagtc tcctgtgttc 20160
ttccttgtgt gtcagggaca cacgtgaagg atagcgtgct tcgcggctgg aatcttcaag 20220
gagatgccat tcacttttтt acctcactaa cacagtgccg tttacaaaaa agattaatgt 20280
acttttcctg aattgactta ctgactgggc ctagagaata agatactggt gctgggcagt 20340
ttggcacaag agtagtataa agaatgcagg attggcccag gtgaaggcat cgtcctaagg 20400
gtagaatggg agtcggtggt tcctggccga cctagcaggt gtactgtggg aagtgctgga 20460
gtgaatcggc tctctgggga gaataagctc atcacagcag ggcttcccga ggagaacgtt 20520
gctgctttga tttctgttgg ctctgaggca gcagcaggtc aaatagttgg ttctctgttt 20580
agagacatct cttgaaacac ttttcgtttt gaccactaga tggtgggata atgttatcat 20640
tttacatttc tgaagaaaaa tagaaatcta actggaagct tttttgtctg ttcagtagat 20700
tttggttgga cccctggtaa acatgggttt cagtgtagca gctttaatgt gttaccacgt 20760
gtgctaaagc atagctgttg gcatgcagaa cggcattacc agcagtaagt gccacttact 20820
tcttcatagt gagtgatgat agttacaccc aggtagatga aattcaggga gagcatctct 20880
gtgcacctta catcttatca ctctgaagga tatgtggttg ggaagcttct cccaaaggaa 20940
cagaacacat cttccacaac tgtataacct atgtcaggca cacgttttcc tgggttgaat 21000
caagcccttc cttaaactgc taacttaaag aatacttact ggttttgtaa agtttggcaa 21060
atgatcttct ctgctcctcg gttttctgtg ttgtgcaata ggaggcaatg gtagtggctt 21120
ttccagcacg gttggtgtga ggcttctcat gagctgggtg acctttgtcc tgatgatggt 21180
ggtgatttta atactgtgta tttgataaca cgattatcta gggtctcctc tacgtctttc 21240
gtccagatgc atctcagcca ccccccttтt gctgttccct taggcataat agtggtaaat 21300
cggtgacatt ttgcttgagt aagaagaagc tgctaaaaac ttctcatgct taaaattggt 21360
aattaagggg acttttтaaa aagaagcaca gttaaaaaac atttccttcc tcgttctctt 21420
ccacccgcct ccctttccca tcacttttat tagatacagc attctgctca ccccattatt 21480
gcaggctcag atagttggtt tgtttttttа aaatcagctt tataaaaaca tttacataaa 21540
ataaaatgga cccattttaa gtgtacattc acggattttt tgtgtatacc tgtgtcacca 21600
ccacaaccaa aatacagagc attttcatca ccccaaaatc tccttcgtgt ccatttgctg 21660
tcggcctccc tgccccctcc tcccacccca gggcagccac agatctggtt tctgtcatta 21720
aagattagtg tcaccaattc tggggcttca gatcagtgga atcatccagc gtgtactatt 21780
ttgtgcctga catcactgaa ggtgatgttt ttgcgatctg tccgtgttgt ttgtagcagt 21840
ggtttcactt ccttttatag ctgagtagta ttctattgta ggcatgtagc ttggtgccac 21900
cagttgatgg agattgggct agtttgccat tttaggttat tatgaataàa gttacaatgg 21960
acatttacat ttgtgtcttt gtatgctttc atttctcttg ggtcattacc caaacttttc 22020
caaggtggtt atggcactgt atattcccac cagcagtgtt cctttcactc cacgtcttca 22080
ccaatagttg aaatttatcc atcttttgaa ttttagccat tcaagcagat gtgtagtggt 22140
atttcatggt ttttttttтc ccaacattgt tttaagatct aattcatatg ctacacaatt 22200
tgtccaatta aagtatacaa ttcagtggtt ttaaatatac agtcaggtat tgcttgacga 22260
cagggatgcc ttctgagaaa cgaataggtg attttgttgt tgtggagaca tcacagtgtg 22320
tattaacaca cacctgcatg acatagctac tgcacaccta ggctctgtgg cacaacctgt 22380
tgctcctagg cataaacctc tacagcatgt gcagttgtga aacagtggta agtatttgtg 22440
tctctgaaat acttaaacat agaaaaggta gagtaaaaat atggtataaa agataaaata 22500
tggtacacct acatagggcg tttactatga attgagctcg ctagactgga agttgctgtg 22560
gttgagtcgt tgagtgagtg gtgagcgaat gtgaaggcct aggacattac tactatacag 22620
tactatggac tttatacacg tcatacagtt aggttacact ggatgtatat tttttggagc 22680
aactgtatta actgatacta taacgttttt ttaaagacaa ggtcttgctt tgtctcccag 22740

gctggagtga agtggcacat ttatggctca ctgtagcctc aacctcctag gctcaagcaa 22800
tcctcctgcc tcagcttcct gaggagctgg gactacaggc gtgtgccact atgcctgggt 22860
aatttatttt tatttttatt tttgtagaga cggcattctt gctacgttgc ccccactagt 22920
ctccaactcc tgacctcaaa cagtcctcct acctccgcct cccaaaatgt tgggattaca 22980
catgggagtt attgcacccg gctcctccca taagtaaata atctatctct ctgttacttg 23040
tggtgggagg aaaagaaaaa aaacacctag gttatgtata atacctaata cgagtacttc 23100
gtaagtagtt attatactgt ttttttttt tgaaacggtg tcgctctgtc gcccaactgg 23160
agtgcagtgg cgtgatctcg gctcactgca acctctgcct cccaggttca agcgattctc 23220
ctgactcagc ctcctgagta gctggaatta caggcacgca ccaccacgcc cggctaattt 23280
ttgcattttt agtagagacg ggtttcccca tgttagcctg gatggccttg aaccgctgac 23340
ctcccgcctc aactcccaaa gtgctgagat tacaggtgtg agccaccacg cctcgcctat 23400
actgtatttt tttttaattt ggccttacta tagcttttt acatgataaa ctttgtaatt 23460
ttttaaattt ttttactctt ttgtaatgcc ttaaaataca ttgtacaaca gtataaaaat 23520
accttatatc tttatcagct ttttctatgt tttaatttta atttttactt ttaaacttaa 23580
aactaggaca caaagacaca cattagcctg ggcctacaca gggttaggaa catcagtatg 23640
tcgctaggcg ataggaattt ttcagctcca ttataatctt atgtgatcac tgttgtgtat 23700
gtggtctgtc attgaccaaa aggttgttat gcggcatata actggattca cagagttgtg 23760
caaccgtcac cacaatttaa aaacattttc gtcacctcaa aatgaaactt gcacccctta 23820
gccctatccc ctattctccc gccagccaag gcagcctcta gtagtctact ttctttctct 23880
gtggattttc cttttctgga catttccaat aagcggaatc atatgatata cggccttcat 23940
gtctggcttc tttctcttag cataatgttt tcaaggttca gcatgttgtc atctgtatta 24000
gaatttcatt tctttttatg gtggaatcat gttccattgt atggacacgt gcgcacgcac 24060
acacacacac acacacacac agaagaacta aatattacaa ggcttatcat gaaaaacaat 24120
ggtctctttc ttgacccttt tcaccctcaa ttcctgttcc ccagaggcag ctcctttcac 24180
acttgtggct gcttctgcag ataagctgtt cggtgacctc catattttaa atactgtggc 24240
cgtattgctg tttcggtttt tcagtttcag gtattatcta gtgactttct gatagggaag 24300
tgagaatttc gtttttaatc cgccccctctg agtgcacctc actcccacat acactcatct 24360
gctgtttgca tggacacatt catgtgcagg ctctttccac tcttgattgc agtgtacatg 24420
atacattttg gttaaatcgg tagtttatgt ttacatcatt atgactgtgg aagttgtgtg 24480
ttaggctgaa tctcagagtg aaccatgaat atatttcctt tcgtggaaaa ctttttgttt 24540
tccctgagct tggcctggtg tcctttgagt ccagagcttc tcaggctcca cttatgtgaa 24600
catggaccca gtgcccccat tggacacagg gtggcagtga gtgggcacag gcaaggagag 24660
aaggagagtc gctccctctt ttcagccttc caccctctgc cctctgcact ttgccccctg 24720
ccccacccca gactgctgtg gcttcacctg cgcctcctgc ccttgagggg ttctgagctc 24780
caggttctga gctccagatg gactcctccc ccgccccagc tgccaggctt gggtttccct 24840
ttttttttt atttgtttga tttcatttcc ccagacagct cttatctact ctttatttt 24900
gttggtttat gtctttttgt tttcctttac tatcatttta ttggggtttt gggggtcaag 24960
agaaaagcat gtgctaagtc caccagattt aaccagaggt caaaaacctt ccatttttat 25020
tgtctaaata ttattcagtt aaggattccc cctccccatc ttagtcccca actgcctttg 25080
ctgaatcttt agcgtctcct gccacagtta ttgcagtatt ccctgactgg cttcctcctc 25140
ctggaccagt gatctgccca cgacccctcc ctcacacctg tccccatgcc ccagacccac 25200
aggacagggt ccaagctcat tagcttagaa agtacaaccc ttggaatcac atgaattctt 25260
tttttgttgc tagtctccta agttgcattc attcactcag tcatacaaat ggtgtatgtt 25320
ttccccacaa tgtcaccctg tttgctgcac tgtgcttgag tctatgctct gcttccagat 25380
ggaagatctg tgtcctccca catctgcctc cttgtcagag ttgagtctgg tgatcatctc 25440
tgacctgaag ctttctctga accatactcg ttatgcaacc tgttgctgct tttctgcctg 25500
gttgtacttc tcttgttaca attactgcac tgtgttcttt tttaaatttg tacatttttg 25560
cagatttctc tgatgcctgg cttaatagaa gacagttgcc ttctcatatc tgcctctgca 25620
ttcagtgtat tggggtggca catgtcgttt tgcttcggaa aattccactg cattgtatac 25680
tgaggggata atgcgagatg agaaaggaaa atcacacgtt agtgttgtta taaagatagt 25740
attgacttta cacaccctca gaaggggggtc agggatgcca ggatgacatt cactacccta 25800
gtgtcactta ccacattgca tagaccatac tgtgccgtac agaggcacat atttctgaaa 25860
cttcctttat tcctaatata ttttgtagaa atttctatat cagtatggat atgtgttttt 25920
tattgcagtg tactttattt tttcaaataa ctgttcgtgt gttagatgtt gaacggtgat 25980
aggcctgtga gggatagttg gagaggtgac tagaggcctt ataaaaacac ttaaacagca 26040
gatgagtgag aatatgctct aaacatggga gtgacagaag gtttttatct aggttgggaa 26100
gaaatttaag attaatattt caggaatgta tgagtgaatt agaagaggag aaacaaatag 26160
tagggcagga gatcatttag aaaatcataa ttatttagac ttgagtgaca gaatgctaag 26220
aaggagataa gggtcacagg aatccagaga tacgaaggtg gacaggagaa atggcaggtg 26280
tgtccacagg gcaggaggag gaggcttggc aatgcggagc attggttgca cacctgggcc 26340
ttggggctga tcgtggtgtc tggacagaaa cacaaaaagg acaacccaat tttggaggaa 26400
agagatgtcc tctgacttca atttctttac gtcccttcta cctctgaatt atctgtttta 26460
tggcctgttt actattaaat gatccattta atagcattta cccttagctt tatgagtacc 26520

atgcactaat aattttgaag tatgctacaa gtcaaaaatt gttgtgtaaa aattgtactt 26580
cctttacctg cctcttgctt ctgttatact taaataccag atagagatga ttttgggaag 26640
tttgatttat actgactttt gtatttgctg ttgtatttat tttttaaaag tctgttaaaa 26700
tgacctagct atggatttct taaattgcta atacatgtgc agatttagtg ctgtgtcaat 26760
gtataataga agcaaatact cattagacta ccttaattta attatacaga tgcaggacag 26820
ctggagcaca cattgatgaa tcattgttcc ctgcagctaa tatgaatgaa cacttatcaa 26880
gcctaattaa aaaaaaagta agtacatgat ttcaatgtag ataatggcaa ttaggaattt 26940
attcgttttt atttttatt tctagaaaat aaaacttcta gaaatatatt caagagttgt 27000
cttaaatatg ctattgatga tattgttctt ttcacatagc atttttaagt gaattacaga 27060
gattatttta tcctatgact tcttcgatag catttgtatg aaatggaaaa gcctgtggtt 27120
ggccatggga agactaaaag gtgccaagag acaagcaaac atttaggtgc tttggtaatt 27180
acttcagaat gaagtttgtt atatctgtag tcaaaatacc tgcattctgt ttagccagat 27240
aaatctcaaa agtccgatgg acctacatcc aagtgtgcaa agtcatttat taggaaaatc 27300
tgctgtacaa atacagttgt ccttcattat ccacagagga tcagttccgg gacccccaca 27360
gataacaaaa tccactgatg ctcaagtccc ttatataaaa tgccatagta tttgcatgta 27420
acctacacaa atcctcccgt atacctaaga aagaattttt tgtagagaca gggtctttct 27480
atgttgccca agctagtctc aaactcctgg ccccaagtga ttctcctgcc tcaacctccc 27540
aattgggatt acaggcgtga ccactgcacc tggctcctcc catgtacttt aagtaatctc 27600
tggattattt aaaataccta atacaatgtg aatgctttgt aaatagttgt tacactgtat 27660
tttttttaa tttgtgttaa attttttttc ttttgaatat ttccaatcgc gactggttga 27720
atccacagat ctggaacttg cagatacaga gggcaaactg tagagttaaa gacattgctt 27780
tcatttgaga tagaattcac attttaacca caacctttc ggctttctat ttatgtaaaa 27840
gttctaattg tgatttcttt atctgagggt actttactct gaaacatcac agccagcttg 27900
ttttcacatg agattctctg ttagagggag gatttgatga ctttctccaa actgaactac 27960
atttcctgta gactagagga gaaataactg tgaacttcac atttcctgaa aatagtcaat 28020
gatatttctt cgttacattt catctcagac aagccatagt ttgcccatgc agtgatagat 28080
gaacttcttc agtcttacct gattataggt gaacaagtgt tcagcagtct ctggactccc 28140
tgtgacatgc taaaatcaag tgtttattgt aaaaacacat cagtagtaca tgcatatttt 28200
ctttgtaaaa catttagtaa acacagactt ctctttgatt gccctccctc aatgtaagca 28260
gctttcaatt tgatgagtat cctaggtggc atttcttcag tacattacac acatgtacac 28320
actcacacat gcatgcttga cgtgaagggg ctctgctatc ttatgtgtat catttggtga 28380
gttgccttct cttccccaat taacaatatg gttttgacca tttcatgtcg gtagctttga 28440
ctctactcag ttttctgtat tgcattatac attgtgactg tttttccggt attcatgtac 28500
ttttagtcac tgtcagtttt tgctaagtat attacttaag ccacatattt gagtttattt 28560
ctccaagtca gatatctaga gataaaatta ctgggtaaga atacatacac attttgattt 28620
taccagctcc accaatcata tacaagatga cctatttctt ggccagatac agtggctcac 28680
acctgtaatc ccagcacttc aggaggccaa ggcgggcaaa tcagttgagg ccaggagttt 28740
gagagcagcc tggccaacat ggcgaaaccc catctctact aaaaatacaa aaattagccc 28800
aacctggtgg tgcacacctg taatcccagc tactcaggag gctgaggcag gagaattgct 28860
tgaacccagg agatggaggt tgcagtgagc ccagatcatg ccactgcact ccagcctggg 28920
cgacagaagg ctctgtctca aaaaaaaaaa aaaaaaaaaa aaaacctatt tcgtgatact 28980
ctgaccaata ttggatgtta ctaatctttt taatttttcc taatctgaag cattaatgat 29040
tgcttgtaca ctttaccact ttaattttca tgtctaaaaa ccttcccttt ccttctcttt 29100
tccaaatgta attgcaaatt aaacccgact caaggcctta ttcttttggg tccttgagat 29160
ggttctgtgc ctctgtctcc cccctcaccc tgtctgctgc ctgcctgccc agcttgctgt 29220
tcctcaagca tgccaatcgt atttctgttt cagagccatt gcgttatctg tttcctctgt 29280
ctggaacatt cttcccccaa aatccttaca cgtgacccgt tttccagcct ccctatggct 29340
ttgtgtagat gttactttct ctgtgagacc tatcctgcca cccgtttata ccagcagttc 29400
cttcccactt gtgccaacta tgcaagtctc ttttcatctg cagtgctgac tggctcctcc 29460
taacacactg tagtaatgtg ggcagtttga tggaatacag ttgctagaga agattcacag 29520
gaccccaaaa taacaatgtg tccaacctgc accgtacctg agaaccagga agcgcaagat 29580
ggagtgtctt cttgtatact gctggccctg agtctatatg aaccagcccc actggcagag 29640
cctccaggca aatccttcat ctcactactc ataaacaatg ttgacaggcc agcacaatct 29700
gtccccaaac ttcccggacc tgtggctata aagcaccact gtctaattag tacattttgt 29760
gtcatgcagg tactttagtg aaagcagtgc aggccggttc caagcctgtt gaaatgaacc 29820
tcccaagaca catacaattt acttatttat tatgtttatt tgctgtcatt ccttaccagc 29880
atacaagctc catgatgaca aggatctttc taggttgcaa gaccagcgcc tgacataaag 29940
tcatgttttt tgtcaataaa tgagtgaata actaacagag caagatcccc agtataggca 30000
ttagccttga gtagctaaaa gaagttcttt ctatgagact ggagcaaaag aagttagcgt 30060
ttacgtgggt agctagctac ccatgtaagc aaatttgggc tggtagcttc gcgctgaaaa 30120
ccaaggaacc tagacagatg acttaaattt ccctggggtc ctataagaaa gaagtcaggc 30180
ataaaagtgt tataggtaaa atcgatgtga agttcagtat gtgtatttgt gctgatggct 30240
gggctaaaga cgggaagtca atgggcagtt ccaagaacag aaagtggggt gggtaaggct 30300

gggaacgtga ggtgtgtttc aaaggaaaca tttcccctgt ctgaggatgg ttaagagtag 30360
agttaaccca agaccttcct gtggatatca gcctggggtt tcatgtgttt gtgagtgtag 30420
ttacagtttt tgggttttac tggctgattg gagttactgt gatttaatga cggtagggca 30480
agcataatca tggttctttt ctttggtaat tataaaatag aaattgtttt attactgtgt 30540
cgtggtcttg cagggaggat gacgtgagaa tagtgctacc aagcaggcag tgggcgtgct 30600
gccaacccac atagagtcca agatcatgcc acttgttttg agaaaagaaa ggctttattg 30660
caagttgcct ggcaaggaga caggaggaaa ctctcaaatc cgcctccctg aggtgggggc 30720
tcaggcagtt tcataggcag agaaaacaaa gtgtgatctg attggatctt gcaatggggt 30780
gatgctggga ggtgtcatct gactgggttg tgtcacaagg tgatgccagg gctcaatctg 30840
attggatcat ggattatgcc atcaggtgtt tactccttaa tttggccccc gttccttggt 30900
ctaagtgctt aggttctgcc cgtggttaca tgcttggttc acctgggcat gctcaagtga 30960
cgtaacttgc aacttcaggg gccgtggcaa ttaaacagtt caccattttg atacacaaag 31020
ttgaactaga ttgggctggt ttggtggtaa gaacagcaaa aaatcgaaag agactggcta 31080
aaaactttca tggaaactaa gaatgctagg atcatgaaaa tgtctcacaa agcataatac 31140
agagcctttt atacagtctt ttaaattctg tccattttct ttataactgc acaaaaaaat 31200
aaatattgcc agttcacata cagtgcaaga aacacctctt ttagaatttt ttattactga 31260
tgttataaaa ggtatcagaa atgtatgcga aagggctttt tctcctgcct taagcagttg 31320
cagtacagca ttaatttttg tgttcttttt gcacagcgta aatgtatgca gcccaaagat 31380
tttaatttta aaacaccaga aaatgataag agatttcaga agaaatttga gaaaatggct 31440
aaagagctac aaaggcaaaa aacaaatcta ggtaagctaa gaaatataat acagttcttt 31500
gcatttgtgt ccatacacct tgtttaattt gcatgatgac tagtggggtt cagcatgaga 31560
gagctgatga agactatgat agctttactc tatgaaggag aaaacaaaat gtcaggagcc 31620
tgcgggagac ttggctggga gccataatag agccacgcag cttgagctaa tcgaccacag 31680
tcttaaccat tcatcaaggt ggtcgaactt tttattttcg ggaatgattt cagaagaaaa 31740
gcaaactttg gctaataagc attattgaaa taaataccta tttatttctt ctttatatat 31800
aactttgtat ttttacctaa ttggcatttt tgttttgtta ccctgaatag gcaaatctta 31860
gatgatacat tattttagtg atttgggaaa atactttaga atattatgtt ctataacaag 31920
atgtcttaga aaaaaatata tgtattctta tgtatatata ttgttaaata atatttttat 31980
atataagaat attatgggct gggcacagtg gctcacgcct gtaatcccag cactttggga 32040
ggcagaggcg ggcggatcac gaggtcagga gatagagacc atcctggcta acatgttgaa 32100
accctgtctc tactaaaaat acaaaaaaat tagctgggag tggtggcagg cgcctgtagt 32160
cccagctact tgggaggctg aggcaggaga atggggtgaa cctgggaggc agagcttgca 32220
gtgagccgag actgcaccac tgcactccag cctgggcaac agagtgagac tccaactcaa 32280
aaaaaaaaga atattatgaa acattaagat gctttgtacg tttttggtat ttctgttatg 32340
ccttttcac tgtcgtctaa agtcagtatt tcctactaat tctgacacag cattgctaca 32400
gataagcaat tatggtcact agaaattcct aggaagcatt aattcctcta gtttttgttt 32460
tctttgtttt aatctatgtt actatgtcac agattctcta ttctgtgttt tgaaattatt 32520
caaatagaat tgtcgagatt tattttattt attttttga gatggagtct ttctccatca 32580
ccaggctgga gtgcagtggt gcgatcttgg ctcactacaa cctccacctc ccgggttcaa 32640
gcaattctcc tggctcagcc tcccgagaag ctgggattat aggggcgtac caccacgccc 32700
agctgatttt tgtattttta gtagaaacag ggtttcacca tgttggccag gatgatctca 32760
aactcttgac ctcgtgatct gcccgcttca gcctcccaaa gtgctgggat tacaggcgtg 32820
accaccgcgc ccggccaaga tttattttaa atctgtgacg ataatgcgac agaactgggt 32880
agaacactta gcccacatag tgctgccaca taattttcca gaaacatggc ctgcatcatt 32940
tgtttcatgc tcagccctcc cgctgcctca cctggtgcgt gtccatcctt ccttcacacc 33000
agctgtctcg tcttcgtcaa agctcaagcc agaaacgtgc aatcgtcctt gacatctcct 33060
tcttcctgac actaaccccc atcaagacca tggccctgct tctgaaatag ttgtttgact 33120
tctttctgttt tctccttccc tcctctctcc cctgatgcct ggatcatccc tcctgcacca 33180
ctgcagccac tccttacgct gccctccact gtctccttac agttcatctc tgtgctgcag 33240
tcacaatggt gaaaacttta aaccagaagg acatcccctc cctggtttaa aatttcctgg 33300
tgtcatccca aggaaaaata ttcaggataa aatcctgtat ttatcatatc ctccaattta 33360
ctaggtgctt tatgatctgg cctctctttc tagcctcata gcaatattgc acactctcct 33420
ataattcttt atacttttgt cactttggcc ttctttccta tgtcagtgac agtgtatttg 33480
aaaatacttt ggcaacatgg taatgataga tacaaaattt tcttcttaga ccaaatatgt 33540
atcgtaatta aaaactatat gtataaagta ttaatgattc aactaatgta catttgtata 33600
ttgtcagaac tacagtaagg gtgattcagg cttaagagtc ccaaaggaga atatattaaa 33660
tgattcttgg tatttttttg ttgggggtga gtatcaaagt tctgaagggc tctttgagca 33720
tatgcaaggt agcattccag aaaaaaacac aactctgcac ccacacaaaa cgagctcata 33780
acttcatggt tccgggacca tgctgatccc acttcatgca gtcaagttca tgtctgggtc 33840
tgtgagtgtg tttgagggta ggagtgatgg ttaatggggg cagtttctga aacctgagac 33900
aagaaacaga aactaaattg cattccagct ttacaacttt taacttctgt gtctcagtct 33960
ttgtcttcaa gtggggatac tgatttgggt ttggatttga ggttggatgc actaatgcat 34020
atattgttct tagcacagtg cttggtgagg gcagttgctc agcagatgtg agccagcagc 34080

gatcgtcaat tgaggagata aatagaggat ttctaattag aagcagaaag aacactggta 37920
ggaattagtg cagttagttc catgttacgc acatacatgt ttgtaatgtg ggagccctag 37980
ttccacttag gatggtaatt tttcatggtc atatcttctt cgtaccaaat ttcttacagt 38040
ttcttcacct agtccccagt ggggctcaag taagtagcag tgatccctga aagtactatg 38100
ttcaaaagtg cttgagatgt tatggaaaat ttatcatgaa agccacagca atgacaaagc 38160
gcaagatggc atcaagatat tagaagtttc aaacaaagcc tcctttcagc gcagggttaa 38220
tccttgtact ctcacctctg tgtgctggaa ttatttaccc atttctctta aacagtctcc 38280
atctttttat tttacacttg ttacatttat ttcctagaag ttggaaacaa gtgataataa 38340
tagctaacat tgatttcatt tttgttgttg taggcactcc tctaagtgtc ttattcactg 38400
ttatctcatt tattctccca ttagccttaa gaggtaggtt ccatcaccat cccattttgc 38460
cagtgaaaaa ccaggacaca gaggtcaaac agcttgtcca aggtcatgtg gtttgtgaat 38520
ggcaaaccca agcttctaac ttaggcagtc tgacatcaca gattacactc ttagtgacat 38580
gtcacattgc ttatcgggtt tttgaaaagt gtgataaaac ataaaacaat tttagatgct 38640
gaataagata tattgagcat ctaaaattaa aagtgacctt atttccaatt actgccttga 38700
agacacctgg ggcacagttg gaagggaagc tttggtggtt acctgtgttc ttccttttta 38760
aagtagaact tcagtgattt cagacagaga gttctaacac ttacgtgacc tccagattga 38820
gtgatttcta caaaacacag gccctccacc agcaagtgct gagcccctat tgagggagcc 38880
agcacgggac tagagacttc ttcatattca ttccagtagc ttatagcaca gtgacgggca 38940
gatgcccacg taaccatggg gcagtatgat gcatgatggt gtgtagcaga gggggcaagg 39000
ccagggagag ctggcaaggg cagtgggagg gtcccaggga tgttgacaac ccaggtgggt 39060
ttggaaggat gaattgtatt tacccagaat aaagtgtgga ggaaagggga aggcccagag 39120
ggtacagagg agtatagaat atttaggagg tagcagcagc ttagcattac tctcaggaaa 39180
tgagtaatcc atataagagt tgaaacatta aagcctacca aatggctcac ttttgaatat 39240
cagtgtaata cgaggacttt agtggaagac agggaaggta agggtgagct gtgttcattg 39300
agggaatgtt tcatgcaagt ctagaacttt ccctagatct tacaacagta gttcttaggt 39360
tttagaatta ttgatctcct ggaaaattta gtgacaaact atggatgctc ttttggaaaa 39420
tgtgcacatg catatggaaa tttgcctaaa atttttagaa gtttgttaca cctcttctct 39480
atccccactg ctatcccata cacccatcaa agcccaggtt ctctagttaa aaatactggc 39540
ctaaaatgta cccttaagtg gaaatgagaa gaactcaagt gtggttaata gtcttcttaa 39600
ctaatagctg tactttaaaa gttgttttat tggtcaactg aaagttgaat atagaataat 39660
ttaaaccact tttaaaagtt agctctccgt taatgttttc cagatgaata ctttgctggt 39720
ggcttacact catcttttga tgatctttgt ggaaactcag gatgtggaaa tcaggaaagg 39780
aagttggaag gatccattaa tgacattaaa agtgatgtgt gtatttcttc acttgtattg 39840
aaagcaaata atattcattc atcaccatct ttcactcacc tcgataaatc aagtcctcag 39900
aaatttctga gtaatctttc aaaggaagaa ataaacttgc aaakaaatat tgcaggtaaa 39960
gtagtcaccc ctsaccaaaa gcaggctgca ggtatgtctc aggagacgtt tgaagagaag 40020
tatcgtttgt ctcctacctt atcttcaaca aaaggccacc ttttgataca ttcaagaccc 40080
aggagttcct cagtaaagag aaaaagagta tcacatggct cccattcacc tccgaaggaa 40140
aaatgcaaga gaaagaggag caccaggaga tctatcatgc cgaggctgca gctgtgcagg 40200
tcggaaggca ggctgcagca cgtggcggga cctgccctgg aggctcttag ctgtggggag 40260
tcttcatatg atgactattt ttcacctgat aatcttaagg aaaggtattc agagaatctt 40320
cctcctgaat ctcagctgcc atcaagccct gctcagttga gctgcagaag tctttctaag 40380
aaggagagaa caagcatatt tgaaatgtct gatttttcct gcgttggcaa aaaaaccaga 40440
acagttgaca ttaccaattt cacagcaaaa accatctcca gtcctcggaa aactggaaat 40500
ggtgaaggcc gtgcaacttc gagttgcgtg acttctgccc ctgaagaagc cctaaggtgt 40560
tgtagacagg ctgggaaaga agacgcatgc ccagagggaa atggcttttc ttacaccatt 40620
gaggaccctg ctcttccaaa aggacatgat gatgatttaa ctcctttgga aggaagcctt 40680
gaagaaatga aagaagcggt tggtctgaaa agcacacaga acaaaggtac cacttccaaa 40740
atatcaaact cctctgaagg cgaagcccag agtgaacatg agccatgttt tatagttgac 40800
tgtaacatgg agacgtctac agaagagaag gaaaacttac ccggaggata cagtggaagt 40860
atgtgaatct ccttttccaa gtcaccttcg ctaaataaac atgtaacagt gcatccatat 40920
tttaaattta tcacaacttt ttcataactt atttccccat ttactcctct ttttacttaa 40980
agaatgtgca tttgatcatt ccaatgataa actctttagg aatagatgac ttgctgtctt 41040
gtggaacttc tagacttatt ggttaagtct gttaggaatc tatttctcca agacttttcc 41100
ttcttatagg tcaaaaggat aagtagtcca tagtatgaat aactgagggg agtgaagtct 41160
ttttccttat tccattggag tcttggcgct gcagcgtgtg taaagatgta tacgatagag 41220
agtattttaa aacctaggtt cttaatagtg aggctattta aagaaagaaa ttaaggtaga 41280
ttaagccatc gattgtatca aagagaaagt gtgaaaaact acttttagaa atctgttgtc 41340
aatattgatt tttgaagaaa ctttggtcag tgttaactat gaagmaacat ttaaacattt 41400
ttgmtcattt gtaacaagcc ttgtttaact tgtacttatt ttgcttgaag catcacttga 41460
aaaggtttac tcctattcat aatttaattg taattataat aaaccatatc attttattaa 41520
aagtcaaaac aataaaaaat tttgcacttc acagttataa gcacaaatag gttccagcaa 41580
ccaaaattga agaaatcttg aactttgacc gtctttacct aaagattagg ttaaaatttg 41640

agtgagaatg cattctctct gcatgatttc tctgctctac aaatgtttta actgcctctt 41700
tgaaggtgga gaagtcatgg tagcgtttga aatcatcaca gacatgttac atacctttc 41760
cttgagtata cgctccccaa aattgtttca caaaaagaat gaaaataatt ttatgttttt 41820
ggcctgctat ttatatcttg gctttctgaa catatattaa atttgacaag aaactgtatt 41880
ttatgttcca ttagccttag tatgtgtttt caaaatattt attttaaaat gttgactcaa 41940
aagttaatat aaaacaatag atgtgtaaaa ttctttggta gttaagaata tcctgttctg 42000
aggtttacat tctccatctt tccagttttc accttgtgta ttttttaaac ttttgaataa 42060
taatgacatg gaaatgtaaa ttaagtagga aaaagctggt agcaaacagt gtggcatggc 42120
ctaaaatccc cgtgttgttg ggagtgtgct agtcctcgga agcaggtgtg ttatgttcta 42180
gaacactgcc cccctgcgtc gacagcctcc ggggttgggg gtaagtagaa gsgggtgagg 42240
ggccagcact agttgactca aggcaccctg gtggggacgg agaggttttt tcgctcagtg 42300
gtgcaggcca tcaggcaggg cccgggtgca agaaaacatt ctgtgtgcgc tagtgcgaga 42360
ggatcttcta cagtcacctg ccttcatgcc attacagaca gacgggaagt cactgggttc 42420
taggacataa aaagacctac atgttggcta gcctaaatcg aacccttttg tagtaataaa 42480
gattcatcaa tgttttaaac tgtcccctgt cagcccctg ggactcaggt gaaccaactc 42540
tctttgggaa tctatcttag aagatgaaac cataaagcct tcagtttcag tgtcagggat 42600
gcacactcta tatctggtga aattatggag gggtgaaaac ttctgtacag caaactgtac 42660
ctccaaatct ttaatgtcga aataaagggc tttttgccat ttctgttttc agttcacttt 42720
tacttgttgc tgttgtcagt atctaagata cagtgtaaaa aaggcttcaa aaacaagtta 42780
caaagagctt caatacgctg atagaacggg aactgagcga gaaacaattt tggttttgtt 42840
ttgttttgtt ttttagtttt ttttgagaca tagtctcgct cttgacgccc aggttggagt 42900
gcagtggcac aatctcagct cactgcaacc tccgcctccc gagttcaagc aagtctctgc 42960
ctccgcctcc cgagtaactg ggattacagg cacccatcac catgcccagc taattttgtt 43020
gtatttttag tagagatggg gtttcacgtc ttggccaggc tggtcttgaa ctcgcgacct 43080
catgatctac cctcctcggc ctcccaaagt gttgggatta caggcgtgag taccgcgccc 43140
ggccaacaat tttgttttct aaaatcttta aaatcattaa ttttttttctt ttttactttt 43200
ttattctctt aattttataa acagtacaca gatacattcc cattgtaaca aagattgcta 43260
agaagactag aatttccatc tcctcacttg cctcttttca ctaattcact tcctaactaa 43320
tgaaagacat gcacccgttg tgtctcaggt gctcttcaag tttgtgggga catagagaat 43380
gaagcagcgt gcaccctcat agaggaagac aaatagtaaa taagtgtata acaatgtcag 43440
ctagcaagca gttaatgata aaaagaaaaa caatactgca ttggatagat aaggtgacca 43500
acgaaggctt ccctgagaag gtgacatccg atcacaggcc tggggaggga gagggagcct 43560
gtgactctgt caaaatccat gtttcagctg gaggtaacag caggaacaaa tgtcctgatg 43620
gaggaaaatg cttgcaggaa caacggggag gccagtacag caaggacttc ctgagctgca 43680
ggaaggaggt tggagagggg taagagccag agctttggga ccttcagtct ctgacaaggc 43740
gggcagctgt tttgttttga agtgtgatga gaagccattg gggcttttga acaggggaac 43800
aaccaaatct gatttaggtt ttaaatgtaa ccatggacac tgaagaacag actgtgggtt 43860
ggagtgtttg tctgcacgaa gcagccactg tccacagttt aatatttcct tccacacatt 43920
tcttgtgtgt gtgtctacaa gcatacaact gcaaatagat attttaagag aatttttgc 43980
atgcatagaa ttatattgcc ttaaaaattg ctttttacaa aagcagtatg tcatatattt 44040
acatattggt accagtaaat cttcattttc taatagagcc tataggtagg gtcagcacac 44100
tttttctgta acagatcaga tagtaagttt attacgcttc atgggcaaag agaccaaatc 44160
gaggtatgta ggtactcatg agatgattac ataatgagaa aaagacattt tccacaaaat 44220
ttttattgac actggaatac attttttttt gtaatacagg tctattaatg agaaaaataa 44280
aataatttgt ggtgggggaa taataacatt tcatttaatt ggagttcaga ctgagtgttc 44340
ccatcaccaa cattgattgc aaatgtttat taaggctgat ttgtaataag atagatttta 44400
cgtatttcac ttttgaaaat atcttttcac acagacagat actcctgatt cgatgtcagt 44460
ccacagttag ataatttgca ttgagcatct tcattgctta gaagacgctg atggaattct 44520
cttagattct tctctcgatg cctgcctctt agcgtgtcct tatattgcag attcatcact 44580
tgcaattgaa aaataggtgg aagctcctca actgtgcagt taaatgggtt ttgaaatagg 44640
aaattccggc caggtgtgat ggctcacgcc tgtaatctca gcactttggg aggccgaggt 44700
aggtggatca cttgaggtca ggagttcaag accaacctga ccaacatagt gaaaccccat 44760
ttctactaaa attacaaaat tagccaggcg tggtggcaca tgcctataat ctcagctact 44820
tgggaggctg aggcaggaga atcacttgaa cccaggagac agaggttgcg gtgagccgag 44880
atcatgccac tgcactccat cctggacaac gagagtgaaa ctccgtctca aaaagaaaaa 44940
aaaaaaaaga aataggaaat tccctttgct cttgcactca gtctgaaaag tgctgctgta 45000
gtttgggctc aggaagtatg tccacagcca gtttgcatgg gaatggagat cttcttgttt 45060
taacctctga cagcacaaga gagaatcgtt gcttatttgt ggaaatgcgt cccacctgac 45120
ccttggcact gccaatcaca gctcttcaac caccgaaagt cagtttgaat tgccaagtag 45180
ttaaagccga ctggtcatcc tgaactagtg cacagcttgg cttctagttg cttttcacga 45240
aggagacaca gttgtcctat aggtgccgtg tgttcactag caaaagcaga aaagtccttc 45300
ctatacccca cttgtccatg ggtttgatac agattttctt ctttgctgtg tgtgatggat 45360
ttttacatgt cagcaccttg tacatacgtg ttgtgagctt atctgagcaa tttggtcatg 45420

tccaactacc agggtcttgt tcatcgataa tagtcaccag ttgttggagg tcaatgatgg 45480
ttaactactc ttctaccttc tatctaccag atcttgttga aggcaagtat cagaaagaca 45540
tttattaaac atttattggc aagcagttag gaagtggtcc acaaattgac caatatgctg 45600
aaggcccagt tctctgtcct ttagtgcagt gtccatactt tatctgaaag gtttgctgga 45660
ggcagacaac attctatggg caagtttctg caaacttgca ctcagcacca gaccatcgtg 45720
tatcctttga ccctgtggtt tattataggg tcatttaggg attaagcctt ggataccacc 45780
tccagggata ccagccacaa ctcatactag atggttatgc tctgttctgt gtgggtattg 45840
ggttcccctg caatatttaa gccaattcag tgttcttgaa tccatgaatt taaccaataa 45900
gaaactgttt ctcacatcca ttatgctgat taacaagctg atgatgtcac caataaccac 45960
tcatttttgt catccatttt ggcttttaac aaagcatcta atattgggct ggaggattta 46020
caggagttgg ggtttttttgt tgttgttgtt ttgagatagc gtctcactct gtcacccaaa 46080
ttggagtgca gtgacatgat cgcagctcaa tgcagcctca acttactggg ctcaagtgat 46140
cctcccacct cagcatcctg agtagctggg actacagacg caggccacca cactcggcta 46200
cgttccccag gctggtctcc aacttctgag ctcatgcaat ctgcccgcct ctgcctccca 46260
aagtgctggg attacagttg tgagccactg tgcccagcct atggtatagt acattttgca 46320
aattctgagc attcaagagg aactgtgaat tactattgtt gcaaataaat agatagacat 46380
atattcatta agtatgttaa attgttgcac ttttgactct tcaaataatt cacaagtgta 46440
ttaagaaccc cctttcccat agcctgccag cctaactcac tggggctgca aaactaagca 46500
atcctagcaa cttgatgtgg gttagtcagt cttaacagaa ggctattgac cacttaactg 46560
tttggttgat tcattcattc atttacatat tcatttttta tctgtcagat gtttactccg 46620
tatctactat gtccaatgta taaacagtga gagaggtaag gttaatagaa agctctgtcc 46680
cttgctttaa agaacttagc taagtaggga aggtacagtc aagatagttt acacacaagt 46740
atcaggaaat tcaaaagtca gagcaattac tttcagtggg aattaaaatt gatattggaa 46800
tgacctctac aacgattaca aaggataaaa ttccgcatta tctattgaag agtgtttttg 46860
tttttttcag aatgaacaaa gtgaacttga tattttaata gatgaatatg aatacagtct 46920
cgttagcaga gttttacttg tgtagaaccc gtataacttg catatatacc aaaggtatct 46980
ctggaaagga atttttccta ggtgtctttt aagattcttt ccagtcttaa tattttgcat 47040
actacattgt aaaataattt catattcaaa tttttgaagc ttagaagaca tttctcattg 47100
gataatgtta agtgtatatt tttacatgtt aaaattatgg attattcagc cttcagaagc 47160
cttttcaacc cttgactctt gcatagtgca ttgtaagagt aaatactaat tgtttaaatg 47220
tgttattaat attagcattg ttagtcttaa ttctgtatct tggaagtagg aaagtaggat 47280
gtggaggaaa ataaatgtta aaaataagag ttatttcttc ggccttagct ctagacaaaa 47340
tttgacacaa gccaagtttc tcctacagtc ttttcatcgt ccacttcttc atctctccct 47400
ttcctagtat ttaagttaca tgtgtcctta tactgtcttg ccctggatct ggctccaaag 47460
tgatcatatt agtcattttc ttctcttttc cctcagtatc aatacttttc cttaatcttg 47520
cttatctctg ttgagtagct gaaggttgtg atttaactaa ttcacactga gaggtgagtg 47580
agtgatcatt tactagcttt cattgatgtg tttgcatttt gatggtatta ttaatccaaa 47640
ctaatttcca aatggtgaaa tttcagataa ctgaaagata aaaatgtggg gtctgtcaga 47700
ttcatttccg tatttgatca tttcgtgaaa acgaagtcaa tgaattgtgt gtgtaatgag 47760
gttgggagga aaatgagagg aagatatatg gctttcacag ggaaatgctg tggaccaaat 47820
tgtgtccttt gacccccaca tttatttact gaaggtctaa ccctcaatgg gataacattt 47880
ggatagggtg atctttggaa gataattagg tttagatgag gtcttgaaga tgggggcttc 47940
atgatgagat taggaccatt ataaaaagac cagagaactg gcttcctctc tctctgccat 48000
gtgaagacag caagaaggta gcctccttca agccaggaag aaagccttca ccggaacccg 48060
accatggggg caccgtgatc tcggccttca ggccaccaaa tctgtggtat tttgttatgg 48120
tagccccagc cgaagaagac agacattcat ccaactgggg tgtgttggag gaagagcagc 48180
taaagggtgc atgttcgttg gaatttcttg gagacattca aaatagatgt ccattaggta 48240
gttggatata gccagccata cctcagctgg gaggtctaga caaggtacag agaattaggt 48300
ctcttcagta atggacgact ttatgggaag tgatgaaatc accttgggga gtgagaaggg 48360
agctgatgac aacccatgaa aaaaccacac ttaggagcaa acacgaataa agagtcatcc 48420
aagaagtggg agagtcagga agaggagggt aggtgtttgt ttacagacct cctgccaaaa 48480
gtggagtcca actaatcttt ccacagatgt tttcagaagt actttgcact ctcaactgct 48540
ttgggtttac cgatgtcaat gttaaaaccc actggcaaat tagtgtggca gagtttatga 48600
aatgttttaa ataaacaaat catttactta gatcattttt tgacttcagg atttgtgaaa 48660
ttgtgaaaac atgttaacaa tatcagtctt ttttttttt taatatcagt ctttcttaag 48720
ttttaaaaga ttgtgttgca tttcttagaa ctttatgttt ataaaatgct ttacagcctg 48780
tttcgttgtt cggcaagaac tgaggcaagt ggctattata aaacttttat tgaatacact 48840
aggaagctgc aaatttattc atgactcaat aacagagcac tacgtcccaa attatatctc 48900
tagtccactg cttttccgat tttgacacac tcatgcttca agtaaatatt tgttatttaa 48960
aaaggaaaat aagtgcgtag tagatataat taataattct aattatttttt aatcttaaag 49020
acgataggag attgcattca tgttctaccc cgggggataa agtgggcctg ggagaaaagt 49080
cagtgcaagt caaccataaa agatacctga ggaggtacgg gatcagtcag gatgtgactg 49140
gtttgagtct cgagtggatt cagtattagg gattatggca aagagtgtag gttggtaggt 49200

ttgtggttta gaactggacc ttaaaatctg tccagggccc aggctgcaaa taacaactag 49260
cttgaattca ggaaagtatt aacattttta ttctacatcc tttttcactg agataggacc 49320
ctgtttttga aaagagtgac agtttttacc ttagactctc caaacttagt tatagctggc 49380
tttatagcat tttatctgca aagaagtctt tctcatgtta tatgattttt aatctctgag 49440
ggcactgatg ttaatttcac gttgcattat atttattcat ctgcatctac attgtctatt 49500
gggttgtgag ctccctaagt gtgggactat atcttgtgca ttttgcatct ccagtgggta 49560
gatgattagc tatttgttaa tcattaggta atcaacagtg cagtttggct atcacctgcc 49620
tggcaggttc tagtaccccc taggctgcta cataactttt gcgtcaaagt ttgcattata 49680
ccattgagac catgttatgg tccatgttag ctcctccttc aaaatcccat gtaagtcata 49740
aagtaggcaa actgtttgaa ggaggaggaa gggtgagagt aagaggcacc ctctgaggca 49800
gtagatgagt caaatcaaag tacacatttc acatttcatc gtgggttact taggtctaca 49860
gaggttagca tctaaggaaa ccacatttca cttgaatgag tatccttttg gtttgtgtgt 49920
cttcatggca agacgctggt ctaaggtgga aacttggggg gagtaaaatc atcatccatc 49980
atttgtaggt tgaagcctga agctctgtac tgaagactat tttctagaaa atctcaaact 50040
gaccccaaaa cttagattaa ttattgcctc taatatggaa ctgcctactc tgaagagctg 50100
ttctttgtca ttattttaaa atctaagaat ttaagtttga cgagtgcgta aggtatgggt 50160
atacattttc ttacattatc aaatggacgg agttgatgct gtagaacact gtaacctgat 50220
tgttaccgac cattgaatta agtgaattgc ttgggatatt ggaatgtaat aaactgaaag 50280
ttctagatag atctcaaaga gccagatata tacaatttat ttaaaaggcc tataacttcc 50340
tgtttccatt atgcataaat gtgatttttg ttttgcttaa gttgtatttg gtccatgtaa 50400
agttctaact aatttttaat ccccttgggt tttaggtgtt aaaaatagac caacaaggca 50460
tgatgtttta gatgactcat gtgacggctt taaggacctc atcaaacctc atgaggaatt 50520
gaagaaaagt gggagaggca aaaaggtcag tgtgtaaaaa tattatttta aactttcaaa 50580
tgctgataca tcataatgtt cttctctggg tcaatgaaac ataaaccagt ctatctgact 50640
tgtcttttat tttaaaaaat tgattatggg taaatgctgg aaaactcaga atatgaaact 50700
gaaagcgttg tttgcattcc agacaaagag ttattattga tagagcaagc tttctcatat 50760
cactttgcta atgcatttct tataaaaatg cctgtagctt ctctcaagca gagaatgttg 50820
gttgtgccag tgtttcttgc cattttataa tcggaataaa tatttactag gtaggaggtg 50880
aagaatccaa acattcattc acttttgaac taaccaagtc ttgacctcaa gccatcagag 50940
tgaaaggttt atatactaac actcaggtac acccttcact ttgtggtttt ggctttaaaa 51000
ccttgctctt cctctgaaag actccgctga tcctcttaca tgagtaatag aatgaggatt 51060
ttaaatgttt ttatcattca atatctactt gcattgctta aatttaaaat tagccatata 51120
tattatacct tgtgcctcat ttttatgagg ccaaaaaagt ataatgtagt gaaacctgaa 51180
ttcagaatgg tagggaaaaa ccataccgat tgaaaagcaa cagatgaaaa gaatgacaga 51240
gtagatgggt ctgcatgggg cttccaggtc ctgatacgca ggcttgaaca gatgggcggc 51300
tgcatttgac ctgcggaaga gaaacctgac tcctttgctt cttatcttgg caatggttaa 51360
aagacattta aaattacaca gatttcatga aagttggcag taacttgtag aaacttagat 51420
ttctttattg atgctttctg gtttgtctcg gaaaaaaaag tggagcaaga aaatggaaag 51480
gaaccctatt tcaggtaaag caacagatgt ggagagagag agactgtcag ggtcccataa 51540
catgtttgtg gcgtgggcaa caccaaggca cctgctctac aatggcgttg cgcactgtga 51600
ctccactgca gcctgcggga cctgctcagc gcgctgcctc ccaggggtgg ggcccttcct 51660
agaacgctcg caacactgtg gctgagtttg tgttttgcgt cccagtttct cagtcttctt 51720
cctactgcta catggccgct tgacctagtt catttggaaa gaaataaaga accagtttcc 51780
tttgcatcta ctaccgttcc cgtgcctctc ctgctgatgc gtcgcatggc accacagctc 51840
tgttctgtgc cctcccgctt tactgaccct ttaccctctg ccagtgtctg cccagggaag 51900
ccgtggtacc tctcatctct attggtactc tacgttgtac catgtctggc ttttttttt 51960
ttaagtgctc agtaaatatt gagtgttgag ttacttgtta ctcaccataa aaatactccg 52020
tcctgtctga tcaaaaggca tgaggtttga ctttctcatt tgcccacagt ggaagttact 52080
gtttcagacg agtggtattg ccttcctgtg cctgggatag ccctgaatct gatgggctgg 52140
gtctgtggaa gcactgggtt agggacaggc atcctgggcg ggagtgtggc cccttcttcc 52200
ttatgaggca tctcactgta aatggcatat gaatgggaga tgggtacctg tttgactttc 52260
tggcattctt ctgtagatca aatagtaagt gctccataaa tataaggtgg tattactgtc 52320
ttgagtaatg ataaaagaat gagtggtcag agagggagac aaaatacaca attacaaata 52380
cacacctcca tatctgcctt caactgctgt gctcaggaac aaaaatattt tcatatatta 52440
aactgcctaa cttgctcaaa tttaagtctt cttttaaaaa tattttaaga gtattagtaa 52500
actttgccct cataatttag aatgtcattt ctgaaacgaa tccaccactt ctggttctgt 52560
gtgaagaatc actcaaagca ggttttaaat gcagattttc tgggccagtc atggtggctc 52620
atgcctataa tcccggtact ttggggcggg cggatcactt gaggtcagga gttcgagacc 52680
agcctggcca acgtggcaaa accctggcca acatggcaaa atcccgtctc tacaaaaaac 52740
acaaaaattg gccaggcctg gtggtgggca cctgtaatcc cagctgctca agagactgag 52800
gtgggagaat cacctgaacc caggaagggg aggttgcagt gagtcgagat catgccactg 52860
cactccagcc tgggcgacag agtgagactc tgtctcaaaa ataaataaat aaatgctgat 52920
tttctggccc cacctgagac cctcctggcc agcagctccc gaccccagtg cggcaccccg 52980

tccttaacgt ggaggggacg aacacctagt gagggcgaag aatccacctt ctgtattgcg 53040
tctcgccaat agcagaagga gcaagaccta ggtttcccct ctttcacagg attttcttcc 53100
taatccagtc cttattagtg ttcaccgcac agcctttgct tgaatgaatc aaaaactcct 53160
aatgccctag ggtagtgctt cctgactggg ctgcgcattg gactcacctg gggatctgta 53220
aggtttgtgg ctgcctggcc ccaagccaga catgctggtg tcattaatat ggggtgcacc 53280
ctggccacta ggatttttt aaactcctga ggtgattcta atgcaaagca gagtttggaa 53340
actactgcct tgggactttt agaatttaaa caagtaattt atcctagaag aagtttcatt 53400
tctttctaaa catttctcat gtaaagttgt ttcattttta gactctaaaa ttaaagacca 53460
aggcttaaag tcctgatttg cgggctgggt gcggtggctc acacctgtaa tcccagcgct 53520
ttgggaggct gaggtgggca gatcatgagg tcaggagatc aagaccatcc tggctaagac 53580
ggtgaaaccc cgtctctact agaaatacaa aaaattagct gggcgtagtg gcgggcgcct 53640
gtagtcccag ctcctcggaa ggctgaggca agagaatggc atgaacccgg gaggcggaga 53700
ttgcagtgag ctgagatcgt gccactgcat tccagcctgg gcaacagagt gagactcctt 53760
ctcaaaaaaa aaaaaaagaa aaaaaaaaat tcctgatttg tttgcttaaa ggttgagtga 53820
gtgttttagg agcgcaaatt tgatagcaat atagatgaag gacgtgtttt attattttac 53880
aggttagaag gaagaatgat ataaatttct taaaaggtaa cattaaattt attttatttt 53940
attttatttt tctgagatgg agtatcactc tgatgcccag gctagagtgt actggtgtta 54000
tctcggctca ctgcaacctc cgcctcctga atttaagcga ttctcctgcc ccagcctcct 54060
tagtagctgg aaccacaggc acccgccagc acgcctggct aatttttttaa gtttttgta 54120
gagatgggtt tcaccatgtt gaccaggctg gtctcgaact cctgacctca agtgatctgc 54180
cttccttggc cctcccaaag tgctggaatt acaggcgtga gccacagcac ctagccagca 54240
acattaaatt ttaagtatat aacttcccag tagtttgaga tcttttgata tgagcatggg 54300
gagagaagtt tatgttgata tgtggtaatg agtccacaga aacactaaaa tttagtttcc 54360
tggttttaaa agtatacagt ggaattgtgg aaggattgaa ttggtgaatt aaaattagaa 54420
gcttctgagt agcagcctac aaatataatg ttagtatctc aaccattctt ttttcccat 54480
taaataggtt ttacctgctt attttgttcc ttgttagatt tcaagataaa ctgtgttaaa 54540
ctgaaatttg gaacttaaca cggcctttt tgtttgtttg tttgagatgg agtctcgctg 54600
tgtcacccag gctggagtgc aatggcacag tcttggctca ctgcaacctc tgcctcccgg 54660
gttcaagcga ttctgctgcc tcagcctccc aggtagttgg gactacaggt gcacgccaca 54720
tatttttatg tataaggaca tattaaggta ttagattcta ttaagcacaa aattgtttct 54780
atttcctaaa gaaaacaaaa tcttgtaatt gaatattaat gttgaaaaag ggagagttta 54840
caggaaatat ctttcaccag ctaatgactg aagcaatgcc tctactagaa tggagaacag 54900
taaggtctgg gcctgacatt tttatgtttt cacttgagag ccagcctaca tgctatttct 54960
gtagtgagga aaatgatttg aaactcagat gtgtcccgtg gccctaatga ctttattttc 55020
tttttagttt taaatctgaa gtagcacttg caggtaatgt cctatctggg cagccctgca 55080
gacaggactg tcagtcgatg agagctgtca gtcgtgagtt ctgagtaatg tgaaggtgcc 55140
aggtagaagg tacaaaggca agaaaggtgg gaaggcctgg agcctgtgcg aagagcagca 55200
cggccttggt gtggcccggg gatggatgca gaaccgcgag aagagagagg ctgacttcag 55260
ccacggccac gggctctggg gttagactgc tctcatcttt ggttttctgt aggttcattg 55320
tgattgttgt accagagtat tgtttttgtt gtttatttac ttgagagtca caggccgtcc 55380
tgtctttgat ctgttctgga aacttctcca ctgtgatttc ttttgcctgt tttctcacgc 55440
ctccattgct gggaacgcaa tctcgtgtgc tatcctttgc ttctatagcc catgtctcat 55500
gattttcctc tatttctttc ctcttgtatc tccttatttc attctggatg tcttctattg 55560
gtttcttttc catttcacct ttgactcttt ttaagtctat tctgatgcta aatccatata 55620
ctgagtttta acatgtatta tttttcagtc cctgctattc catttatttt ttttaattat 55680
tttttgtaga gatgggggtc tctccacgtt ggccaggctg gtctcgaacg cctggtctca 55740
aacaatcctc ctacctcgtc ctcccagggt actgggatta caggcgggag ccttcatgcc 55800
ctctatttga tttataaaaa ccatttccag ttctctgtca aaattattaa tcctatcttt 55860
tatttatttg aacatattat gcatatttct tttgaaataa ctccctttttc tggctcccct 55920
caatttctgt ttttcttatc tgttgttttc aatcatacgt tccatatcta atatgcctgg 55980
ttagtttgtc tttatcttcc tagcagggac tgagatgatc tggagctggg gttctgtctc 56040
tgtgaggcta gctgtcccct gggagtgtgg gcttctgact ctggttcacc tcctcttcca 56100
tgggtttctt cttccatgac tcactgattt agtagctggg caacgtctgc aaatagctgg 56160
ggcttgtttg tttgtttgca tcttgtccag cttttctgag ggctcacagt gaggagccta 56220
tttcaaacta cttagtccac cattcctgga gacgatgggt gaattttaac ggccacttaa 56280
cttttctaaa tagagttttg gtgtgaatgc ttctctgaga agacagcagt aagaggccaa 56340
gtcaagagaa atgatttttg agatgaacac gtaggtcagt ttgcaaaaga cacactaaac 56400
acctgaattg acattaattc agtttctctt aaagagtgaa aaaaaccatg attccatgaa 56460
gaattataga atctcagagc tataacttcc attagctttt tttttggtgt aatgcccatt 56520
tttaatggca aaaatcactc tataaatcag ccagaaaaag agtctgtttt tttttagact 56580
tattttaaat atacttgttt caaatttgtt gagacttttt ttttttttt ttttgagatg 56640
gagtctcgct ctgttgtcca ggccgaagtg cagtggccca gtcttggctc actgcaacct 56700
ccaccccacc aggttcaagt gattcttgtg tctcaacctc ctgagtagct gggattatag 56760

gtacctgcca ccatgcccag ctaatttttc tatttttttt ttttttaatt agtagagaca 56820
gggttttgcc atgttggcca ggctggtttt gaactcctga cctcaagtga tgtgcccgcc 56880
tcagcctccc aaagtgctgg gattacaggc gtgagccacc acacccggcc tggtgagact 56940
ttatttggag gatccagtta agcagtttta ttacctctgt aatcttagtt gcagcatgta 57000
ggtcattgac attgatagtt atacatcttt tcagagggag aaatagaaaa tattatgacg 57060
aattttgacc tgttttcttt gttacttgtt gaatattgtc agacacagaa cccaaagaag 57120
ctatgtatag ataccagcac tctggtagaa atacacgaat gtaatttttt tttctccaag 57180
tatttggttt attctactac ttctggattt ggtttttcaa aatattgatt attatcctca 57240
ggaacatttt taatgtgagt tatcaacagg atagcttttt gtaagtggct cagttgtaga 57300
atctcatttt ggagccatct ctgccaatcc agcttgttgc atgtgaaggc aagctgtggg 57360
tcagagcaca gaaatgttta cagaggcttt cctaagcctg gaggcctgga gagatgtgaa 57420
ggaacaaata gagcatactt attttgatag tggtttaaaa aaattaaaga attacacacc 57480
acatagaatg cttaaattcc tgaaagtttc tcaaataggg tgcaaaacaa ataatagctt 57540
gcatatgctg atagttgctt gttcttacat ctttgctaga atatgagccc ataaggacat 57600
agtctatatc ctgttagtct cttaatactc agcaggatat agcatcacaa acaaaataag 57660
tgctcagtaa atatttttctg agtaaataag agatgcatta atttcccttt tacttttttca 57720
gtgaacatgt ttaaaacatt tttggtgctc ttaaccatca ctcagtaatg atggaatcat 57780
catcatgtac ttcacttatt tttgaatatt cttccaaaac ttgagagact gtcttctttc 57840
agtaaaagat ggattctctt ctccaaggct gtgcatggca gcgcagtgtt gctaaagcat 57900
tgcccccaga gccagatgcc tgggttcagt cccatctctg ttactcacct gctctgtggg 57960
ttccatggtg ttgaacaaat tacttaatat ctgtgcctat acttctttgt gtataaaaca 58020
ggaataataa taatagtacc agtctcctca aagggtttgt gctaattaat tgagttgaaa 58080
catgcaaaga gtttaagata gtacctcata tatagaagtg ctcaaaaaat gttagctatt 58140
ttcttcagca ccagcttggg tgagggtcat gtctgcatat tgactgtgct ttgttctgca 58200
gctataactt ggagtaggtc tctcttacct gcctcctctt tgcccactcc cagagaccac 58260
catgtgtctt taatgaaaat gaccctcaaa actctgggac agtccacact gtgtttcttg 58320
ttggacttac tgaccacagg catgccagag ccaaaataga gtcttgggca gggggtgagt 58380
ataggagtat agccttttct aaaagctcct tcagtgattc tgagctgatg gtcatcctcc 58440
cattgagaac ctttgttttg ggggtgagat gtaggccatt agcatgaaat tgtgctctgt 58500
catctccccc aggaggcaga agactgagtt ctgcggtcag aaatgcccgc ttgggggatc 58560
tgcttcctca gttttcgaga gatgctttcc tcatctccag tatcattaga accttcctga 58620
aagaactgag atctttgtga gctgcgatag ggtactcaca gctgtcattt attgagcatt 58680
gtgacctctt tttagattga gttttctatt tctcagtcat atggaaagct gaaaagaaag 58740
tatatttcag agagctctaa tcatgtcttt attgcggagg cagtagattg ggaattacag 58800
ctcatttggg tgtagcatcc ccggagaagg agccttgcag tggaaagaag ataaaagggt 58860
cccagtggcg ggaataaaaa gagtactaga tgcccagagg gtgggaaagg cctagcccag 58920
atgcagtgtg gccaggccag ctaggggcag gaggaaagag agctgcaggg atacagatgc 58980
cttcctgagc agagaaaata gaatacttga gccaattttc atgtaaaatg gattattttc 59040
ctggcgtttc ctgtccttca agtaaaaggt tctggaatga gtacttcact gctgtaatgg 59100
agacactaat attttatgaa tgcagtttta cagtttgcag taatgccagg cctttggctg 59160
ttttccatta gatggtgcac ttggctggaa gcatatactc ttgtagcttt gattttaaat 59220
ttaactttca agttgaaaga gcagtgactc atccaaagga caggtgatat ttatttattt 59280
tttcttgaaa atgcagcacg ggtatgttgt tatcacacgt ttaggggaat tgccacactt 59340
cctcgaggat gacacccttt gtaaatatcc atgtaaatca tttccattgt tcagacccgc 59400
tgtacgcaga aagataggcc ctttagtgcc gaccagccgg ccagtgagct ctgtaagatc 59460
gaaggtgccc ttggtttcca acacagctgt ttcagtgatc tgtaattgct ttgataaatc 59520
acttttggca gagtgtaccc agagctggca gtggcgggga tgtgctcgtt gtaacaggtg 59580
tgcggtccat cagcagatgt tgcttgatga agccatttaa aaaacagctg cctgttgata 59640
gcctaacagt tgctttcagc ccccattagc acgttgtttt tttcttgtta tgtatgagag 59700
aaaatatttc tacagaaaac attaaatagg atcttcaaag aactccatct ttttaaaaat 59760
gtgttttatt tgttcactaa ctgattttgc atgcattgta aatgtgtggt tcagaaattg 59820
tcaaatgtgt tttggactgg acgtggtaga aatgaggacc agccagggtg gatctcctgt 59880
gcctcagtgg tcgtctttgg ccacgtaaag gtagaggcca ccgacggagg acatttccca 59940
ctgggagacc cacaggcgct aagagaggag ctagccgaag aagtctattt aagatctgct 60000
gctttggcca ggtgtggtgg ctcacgccta taatcccagc actttgggag gccaaggcag 60060
gtggatcacc tgaggtcagg agtttgagac cagcctggcc aacatgggaa aaccctgtgt 60120
ctactaaaaa tacaaaaaat tacctgggtg tggtggtaca cacctgtagt cccagctact 60180
cgggagacta aggcaggaca atcacttgaa cccaggaggt agaggttgca gtgagccaag 60240
atcatgccac ggcattctgg cctgggcaac agagaagatt ccatctcaga aaaaaaaaaa 60300
aagaaaaatt ctgctggtag gcattctatg cactgagcaa aggagagatg tggaggccca 60360
atttaaatag ttacagctgc tagctcctaa ggtctatctt actatctgca ccgtttgcgg 60420
ggagtcagct taatgatagt aaactgtgct aaatgggtct agaaatatcc aattaatctg 60480
tttgagatat tcggaaactc aatagcttgc tgaagtagca aacttgaatc cttatttta 60540

ttttaaaagg gagtaaaggg actgtagata agtaaaagat gctctgcact gcgcctctct 60600
ggtaccagtc cctctcgttt aggcagcggc cacttcccgc ggagctgttc acgccaagtg 60660
accctgccac tgcgctgctc ccaccacccc atgtccaccc cgtcctcgga cgcctggtct 60720
cagcacatca ccggtattct cttcctctta ccagtaatta gtttgagact gtgactcact 60780
tctgtccaac aagatgtgaa gggaagtctt cctgggaggt ttctggaaag cgttctctca 60840
cttgtgatag ccctgggaag aaatgctccc cgggtcctca gagctttgtt gtggctggac 60900
gcatcttctg gaactgcgac agcggaggag gaagccaaga gagtgaacca aaacaaggaa 60960
gggcggaggg cgggggaggc ctgcaaacct tacggcttat ttccactgac atcagagact 61020
catgttaata agtaacaagc ggctttgttt gttatgctcc tcagacacgc ggtaagggag 61080
acacacagaa atgcacagct gtacgtattt gtcttgaagg ctagaattta ctttaaatgt 61140
gagtggtttt cccaggaaaa atttatgtct gttctcttga ggaataatta tttcctactc 61200
aattttatct atcgatccat ccatccatcc atccatccat ccatccatcc atccatccat 61260
ccatccgata cagagcctcg ctctgtcgcc caggctggag tgcagtggcg ctatcttggc 61320
tcactgcaac ctctgcctcc ccagttcaag tgattcttgt gcctcagcct cccgagtagc 61380
tgggactaca ggcccgtgcc actacacctg gctaattttt gtatttttt ttttttttt 61440
ttttcctgag acagatcttg ctctatcgcc aggctggagt gcagttgcgc aatctttgct 61500
cattgcaacc tccgcttccc aggttcaagt gattctcctg cctcagcctc ctgagtagct 61560
ggtactagag gcacgttcca tcacgcctgg ctaatttttt ttttttttga gatggagtct 61620
tggagtctcg ctctgttgct gaggctggag tgcagtggtg ccatctcggc tcactgcaac 61680
ctccacctcc tgggttcaag tgattctcct gcctcaacct cctgggtagc tgggagtaca 61740
ggcgcgtgcc accacacctg gctaagtttt tgtattttcg gtagcaacga ggtttcgccg 61800
tattagccag gatggtctca ctctcctgac ctcgtgatcc gcccgccttg gtctcccaaa 61860
gtgctgggat tacaggcatg agccaccacg cgcagccttt ttttgtgttt tagtagagac 61920
agggtttcac cgtgttggcc aggatggtcc gatctcctga cctcgtgatt ctctcacctc 61980
ggcctgtcaa agtgctggga ttacaggcgg cagccaccgc gcctggccta atttttgtac 62040
ttttaagtac agacggggtt tcaccatgtt gtccaggttg gtctcaaact cctgacctca 62100
agtgttccgc ccaccttggc cttccaaagt gctgggatta cagggttgag ccaacgcgcc 62160
ctgccctcaa ttatatttat ttctttgcct ttccttacgt ctttaactct tcacactttt 62220
aaaaaagtta ttgccttcca aataatattt aggaatataa attatttgat attaatccag 62280
ggtaatttcg atttgttttt aaaaaagggg aataaaaaca ttattattca gaaggggtta 62340
aatacaatga caaaaactgc aattcagaat taatgaggcg ttataatagg gtttgttaaa 62400
aaaattatga ggtatttaaa atagattttt ggcatatcct tttgtgactt ttggatagac 62460
ttaagactta gtttatatat caatagtgag tctgtatagg aaaagaatat aatattcagt 62520
gactgtcaaa ccagtgactg gagcagcttg gtatgaagcg cttcttattc tggtctccct 62580
aatcagtgat tttcaatttt gaaaactttt ttttgaagtt gtgttgtttt atttttctgc 62640
agaaatatct tctgcttttc attttaaagt atatttgcta tttatttgca atctagttct 62700
catcattaaa agcagtacta aaatcttatc ccagaattta taggttgtgt cttttgtcct 62760
ttttttgttt ttagtatttt tctgtcactt tacttcctca ggtgaagttt taacaaaaac 62820
gagggaccat ggataggaaa gtaggaatga aacagtttac agggttgaag ttgtggtata 62880
attctttttt tttgttttgt ttaaagacag ggtcttgctc tgttgcccag gctggagtgc 62940
cgtggcgaga tcatagctca ctgcagcctt gattgcctgg gctcaagtga tccctccagc 63000
cttggcctca tgagtagctg agactccagg caggtgccac catgctcagc taatttttt 63060
tgtttgtttt agagatggga tttggctgtg ttgaccaggc tggtcttgaa ctcttggcct 63120
caaaccatcc actcgcctgg gtctcccaaa gtgctgggat tataggcatg aaccaccatg 63180
cctggcccat ggagtaattc ttgtggagtt ggaaggtaga ggtgtgtacg tgtctgtttc 63240
tcaaaatagt agcactagcc aggaaatcca tgaatttgca tatttttccc caagttcagc 63300
ccatttgctt tggtgagttt ggggttatac ttagagtggg tagtataagg agtttctgcc 63360
ctacacctta gcttaagcaa tttgagcaca ttgctttttg agttcaccac caaggatcca 63420
gagctcagag gcagtctttc ctgtgcagat aagagtgcac cctgcctgca cctcacggtc 63480
ttgggctctg tggcttctct cctcctgcca ctgcccctta ttgtgggtag gctggaattc 63540
cctatggtcc tttgtttggg gaaggggat gcttggatgt tcccgggtgt cacctgtgca 63600
tgccccctat gctgtcctcc cacctgccct gtcctacaag catgacctgc acccttctcc 63660
cacacaccca gaccgcagct tattcttact ctccctggcc agcccctctt cttggagagg 63720
agaaaggatg atgtgaaaat aatatctaac attggggctc cccagcgact tccacaagga 63780
gcaaggagct aggtgcatgt gtagacccca tgggagcttt agtgttagat accgagtttg 63840
ctagatgaaa catcttttta attgaggtgg tgcagatgta ttgtttgaac actttagaca 63900
ctaatgatga actacttgga tgtacatttt tttggttttt ttttttttt gctatgaaaa 63960
ttagaaaaaa tatttatcca agacagtaag tattgaaaac tgatactggt gctgtatgga 64020
tcactattat tgtattattt gaaactgttt ggaaaaggta ttgtagtttt tagaaaaaca 64080
aagcaacctg aatattaaaa gtctgtgaat ttgagtaaaa aacagtccac ataagggaaa 64140
aaatatataa ggaaggacaa tgaagttttg aaactgttac tataagaaag ctaaaggctg 64200
agcacagtgg ctcatgcttg taatcccagc aatttgggag gctgaggcag gaggatcgct 64260
tgaggccagg agttcaagac cagcctgggc aaaggagtga gacctcatct ctactaaaaa 64320

taatttttta aaaatattag ttggacatga tggtggccac ctgtggtcca agctactagg 64380
gaggcttgag accaggaatt cgaggctgct ctgagccgtg attgtaccac tgcactccag 64440
cctgggcaag agtgagaccc tgtctcaaaa ataaacaaaa aagaaactta aagattttag 64500
tctcaatttt ctacattgaa cccatcttta gatcatagca tgtataaaat taaaaatggg 64560
ggaatatcaa cattattata tttaatgcta tagcttatta ttgtatttaa taagctactt 64620
gtttaaagat ctggggtctc ttgggtccac agactgagtc tttctgaagg tgctttacac 64680
gatgtagctg ccagggatct aggtcatata atatcctcag gatgggattt gaagacattt 64740
ttccagaatt tatcttttgt catattggat tttatttta aaaatttcct ctatagtcaa 64800
aatttatata aatatatgat tctgatagta ccatatatat ttagatgggc ttatactggg 64860
cgtgaacaag gttaataatc tttgtgaata tgtgggttat ctccttattt tacttattct 64920
taaggaaaat taatttcact gtttaccaaa gaactgatag ctaaacccaa aagatttcaa 64980
agaatgtttt gtttttgaaa tgtttctatt tatcactaat aaaacgggta tatctgttta 65040
agttgaccta tctttggtct tactaaaaca aaatcagcta gaccatttcc caaataatca 65100
tgcattcaat actctttttc tctctctctc cctgctccct catctctact cctttagaac 65160
tttcagaaca ttcttttgtg tagatacagt gtttcatgtc tgttattgtt tctcactggt 65220
cgttggattc tttcatgtga ccacctttt cacgtttgct ctgattgcct ttggatgcgc 65280
ctaactgtgt gcttttcctg ttaaggaaaa gaatcctgca tgtttttttc tcatcgaata 65340
acaatgttaa aaacagaaaa gggttgtttt tcttctttgc agtaggcatt ctgtagtaga 65400
taccttgaca tacttaaatt tgtgagatgt gtctagacga atggaagagt aatatctcat 65460
attaatatat tgctaataat aagataaagg tttcagcttc ctggagctgt ccatataata 65520
gaatttgtac ttgtttttttc atttctgaga tcctcatact ttggggtttt ttttattttt 65580
ttatttttttc gagacaaagt ctcgctctgt cacccaggct ggagtgcagt ggcgcgatct 65640
ccgctcactg caacttccgt ctcccgggtt caagcgattc tcctacctca gcctcctgag 65700
tagctgggat tacaggtttc ctgccaccac acccagctaa tttttgtatt tttaggagag 65760
ataggtttca ccatgttggc caggctagtc tcgaactcct gacctcaagt gattcgccca 65820
ccttggtctc ccaaagtgct gggattacag atgtgagcca ccatgccagg ctctgagatc 65880
ctcgtacttt taaataaaat gttaagatac atgctttatg cttttgctgc ctctcatgtt 65940
tcatgaatac aagtaaaccc atgagtaact catgaataca cataaacttc tgggcctcca 66000
aacgatgccc tgccagtggc catgccacag gaatcagagg ctgtacttca ctttgtggtt 66060
gctttattat tccaccatta taagctttag tagaaaatgt aaagagggtt gttaaactga 66120
aggagtgttg tctcaaactg aaggagaaaa gtagtgttgg tgctgtaaga tgtacataaa 66180
ctaaggggtg tcttttctac catccagtta gcaattagga aagtccttct ttgctcatac 66240
cattccaaag ggagtcatct tattctttct ctaaatttcc ttacaatgga ggctgctaca 66300
gtttaagtat cgaaggtcct tttttttcag atttcacctg cagtgcctat aaatttgggg 66360
gaatgccttt ttttgggggt gaccaacata ctcagtggat cttggaccta ccaccaagtg 66420
accttccttg ctcacctgta aggctgagaa caccgtaagc aaagtaccag gcttctttcc 66480
ccaagagggc tttgtaagcg ttggcgccat aaaatcaacc tgaggactta ggtggctggt 66540
tatttctgag taagtgaata tcactctcaa atacgacatt ccagcaaagg ccatggttgc 66600
atagccactg tttttagtta tgtcctggta actaggaaga tggattgttt tttaatctat 66660
gcaaataatt atattgcgct gaaaaaaatg atactcaatt acagtttcac aattctggag 66720
ggatcaggca gggataataa gataccattt ccagatgttt cctttctgtt tataaaagca 66780
tagtcgactg aattgttagg agatacaggc agagggagaa gagaaagggt tccttatgta 66840
tccagaatat agagtgttaa aatagcaaca atactgtaaa caaaagccgc agtcctcctt 66900
cagtagttca tctgggccta gtcattaatt tttgttccac ttgatcttgg gttagcagtc 66960
tcatgaatcc gtctgcttct caatgagggt tatagaaatc ctcttccccct ggtggggtct 67020
cagcattatt tagacaatgc cataagaagc ctgtacccaa aagtacccag tatagttctt 67080
ctccacgggg ctctaacaca gccccctctt ggtcgaaggt aagtcactct ggcctatagc 67140
taattgcaga tgctgatcag ggaagtgtca gagaaacaca gaaatctgta ggtgacaaaa 67200
gattttaaat ggctatggtt ctcgtattac tgataatttt caaaactaaa tttattgaga 67260
gttcattaca acagtattgg caactgataa gtaaagttag ttatggtgtg caaaacagag 67320
tcaacccgaa aaagttctag atacaacatc tagaaacacc ataattaacc ttattttaaa 67380
agaacagtgg atgttacatc taatttataa aaatggaaga acataatctt tacagaaaaa 67440
atcttcagat ataacaaaat agtcccaaga catartatac aatgaatatg ccaagcatat 67500
aattagaata gaccaagaat atcacatcaa gagggttatt ttagagggga cataaacacc 67560
tatgtattaa taacatatat ttaacctagg gctggctatc ttttttgatg tgacaatttg 67620
tcccatataa cttatcaata gtaacacatc aaatggatct cctaattatt tcaagcatct 67680
gtttttttatt aaagtaaaag cacaaatact ttttattttc caggtatgtc tggggaatct 67740
tagacagttt tttgttttgt tttgtttttt tgagatggag actcactctg tcacccaggc 67800
tggagtgcat tggcccgatc ttagctcact gcaacctccg cctcctgggt ttcaagccat 67860
tctcctgccc cagcctccca agtagctggg attacaggtg cctgccacca tgcctggcta 67920
atttttgtat tttttagtag agatggggtt tcgccatggt gtccaggctg gtctcgaact 67980
cctgacctca ggtaatccac ccgcctcggc ttcccaaagt gctggaatta cagggataag 68040
ccaccatgtc cagcctcaga cagttttaag tacaaaatat atcatttagg atttgatttg 68100

cggaaggcaa aatatcaaaa attatcaaga aattttgaat acctgattcc aataggatca 68160
tgtaacttag aaacaatttt tgactaccta tttaatcaaa gtgactgtaa aaggttttaa 68220
aagtaaacag agaggtaaca tgattgtaaa gaaccttagc tctttcctaa gagacacgaa 68280
ttcttgaata ctcaagggta aaataaagtc aatataaacc atagaaggtt attctcataa 68340
aacacagaat ctttggaatc taagccaatt atacagaaaa aagaataagc ctttatttt 68400
taggtgaatg tggtaaacag taaaccaaag aaacaggctc atcaatattg ggtaaacttt 68460
tctttgtttt taaatgttta gtctttagtt ttaagagatc atctgcattt tttctgtaat 68520
aaacttaaaa gatatccact tatatttctt cagatttatt aattctgtag cattttaagc 68580
attgaaatga cagtttttct ctcaatcctt ttttttttt ttttttttt tgagacggag 68640
tcaggctctg ttgcccaggc tggagtgcag tggcacgatc ttggctcact gcaagctccg 68700
ttctccccag gttcacgcca ttctcctgcc tcggcctccc aagtagctgg gactataggt 68760
gcccaccacc atgcccggct aatttttgt atttttagta gagatgaggt ttcacagtgt 68820
tagccaggat ggtctcgatc tgctgaactc gtgatctgcc cacctcagcc tcccaaagtg 68880
ctgggattac aggcgtgagc caccgcgccc agcctgtctc aatccttaac aatgctatat 68940
ttgttgtatt tcatatgttt agctttctca tggagaaaaa gaaacatagg cataaacctt 69000
tatactatcc gcctgctggt cctgcaacat gagtttaata aagcgttcct gatacttaaa 69060
caatttctat gatgtcagca gagagatatc agcaagagtg attgtaaagt agctagcctt 69120
ataagtcaag agttataatc tttgatccac tgctcaatcc atttcaagat ctgatctaca 69180
ttattttcta gctcttctgg tttattgctg ggcagccgat gcacaacttc ttccttgtag 69240
gatgccgtgg cttcttcata aagaacttgg aaaatctcac actgaatatt gtcttttagt 69300
ttcttctcat tataacccct catttgaagt atttcgtaca atatgttggc atctattctc 69360
aacacaaaaa ctatgtgaaa ccagcgttca gggaagaaat cacaaccgtg gtaatcaaca 69420
ataactccac attctctcat ttggttatct aactcatcag ctactctgtc ttcatctaaa 69480
atgggacaat tatactcttc atcatagtcg tcatacaatt rcttttctca agctaaatca 69540
cccacattaa tgtatttcaa tcctgatttt gattctgata tgtggttttt ccaacccctg 69600
gtgtacctgg ctttctatga cacgtttcta tcaccaagtc agaacaaagt gacactttag 69660
gactgaactc agggagtctg tggggtcaaa actaatttca taatactact aagactttaa 69720
catgcaatgg gttcaccttg ctgtctccaa aaaaaaaatt gcaccactgc actccagcta 69780
gggcaacaga gcaagaccct gtctctcaaa agtaaataaa taaataattt aaaaaattat 69840
tgttaaaaaa agtttgtcag gttaatgatt caatttgatt aagcacaaat ttacattttt 69900
tcatagtctt aaactttagg agtaacgttc acttatttga tcagtaaatc tgtatagctt 69960
ttgtaagaac atgtaaaagt agaatagcaa tgtatagtgt ggctgggcac agtggctcat 70020
gcctataatc ctagaaattt ttggagtcca agatgggagg attgccgagg gcaggtattt 70080
gagaccagcc ttggtgacat agcgagagac cccatcttaa aaaataagaa taataatact 70140
taatgctgac aactcataga agacatgact atttttatta aaccccaaat attcaactag 70200
tctcatttgc caaatattta cctaaatgtg tgaacttgaa ttcttaaaac atttacgttt 70260
ctataggaat actttttttа gtgctgttga aagtattatt ggaagttcaa tttccttaat 70320
ttctgggaat tttaggaaga ttcaatttat aggtgtctct ttatttctaa gccagtcaga 70380
acagaacatc cttaagagct atcacattct cacttggtaa gaccatctca tgatggttat 70440
cccaggatga gagacaatag ctgctttgaa agttcccctg ccacactggg cttccagtac 70500
cagtgcagct aatgaccctg ccctaacagc aaatgctggg gagcagggtg caagtgttta 70560
cttgggtgcc cttcacgggc actcctttta cgtggtggac agcctgatgc tttgttctct 70620
aaaccagtat caggcattcc tctcatggga gatgtgctta tcctggcaga cgcccttgtg 70680
gctcttttct gacccctctc cagtttatga ctgcctgacc atcgctctgg tgctcagagc 70740
ctgcccttgt gttcctcccc agcatcccgg ggaaaaccca ggtagcctgg gagagcccct 70800
ggttcttcag atggaatgtg caaattcagc acaccaacac gataggaaat aagttccaag 70860
atttattact tccagatcct agagagggag ggcgccatga gtcgggaggg caatgctcta 70920
tccccaggtc accagaagaa tgaatgaagt gtcaggcata gagcaagaga gagtgggacc 70980
catgggccac cacctttact gggggccagg gcattgtcca agcaggtttc ctgcagggag 71040
ttttagttgg tgagtttaaa acaggcagcc atgagtttca ggatcacaca gcaactgaga 71100
ggtggtccct gtggcatact ccacagtcca tgtggggtgt ggggttggca gggcagccag 71160
gtagactgtc tcttagagag gccgtcacca gaaagaggag gtgtataagg cagatccctg 71220
gatcaacccc attgaggact gggggtggca ggtggaagct gtcgagggaa actaagccct 71280
gtttctggta tgagaaggtt aaacttatca tcaaaataga tgccaaggct atatgaaact 71340
gtcagtattc actacagtgg catttccaca gtacaataca gacatacaaa cagacataga 71400
taatttgtaa gctgtaattc taaaatttca ggccaggcgc ggtggctcac ctctgtaatc 71460
ccagcacttt gggaggccga ggtgggtgga tcacctgagg tcaggagttg gagaccagcc 71520
tggccaacat ggtggaaccc tgtctctact agaaatacaa aaattagctc ggtatggtag 71580
tgggcgcctg tgatcccagc tagttgggag gctgaggcat gagaattgct tgaacccggg 71640
agatggaggt tgcagtgagc cgagattgca ccattgcact ccagtctggg caacaagagc 71700
aaaactccat ctcaaaaaaa aaaaaaaaaa agaagaagaa aaaaattcag tcatagacca 71760
aacttaaaag cagaaatata aaatttтact cagatgtcta cttcctgatg gcatgaaatt 71820
cttaattgtt ttgaaaccaa agtagaaaag cagacaaacg aaaaatacta gcaaatcaga 71880

ttctgttatc tttcacccaa cagagacaag atctctataa accagcagtc cttccccaaa 71940
tacgtagtat acaaaccgct tcatgtctgt cattttcgtc aaccctgggg tccttcaaat 72000
gccttttgtt ccttctcatt tacttcacct tgacttttca agacatattg gttatactac 72060
acagttggtt acatttgaag tatttcatgt aaattacaaa agtatatgaa taatgtgaat 72120
tcatttttgt ttatatatgt atatgcatgc atacacatac acacacactc ctatagagtg 72180
aacatttggc tgaatatact gccaaattgt taaacaatag tcatttctag ctggtggaat 72240
tacaggaaaa tttgtgtttc tgattatata tttctatagc atttaaattt tttgcaagtc 72300
agcgtgcatt tcttagataa gcaaaaaaaa aattaaacat tttatttaaa tttttttca 72360
attccagtta atagcagatg tcaatagaac aaataagttc ccttatccat gcttctgtat 72420
gtgggggatt cacttgacag gtgcaacaga agcacaagca ttattgtgca cctgtgtctg 72480
aaatgagaat gaggctgcct agaagtcttg agaaaagtgg ctgacgagtc tacaaaaaca 72540
cccttcttac cctttctcac tttgaagtgc atgaagacgt tgacacactt ggaggtctgc 72600
tggctaactg gtggaacaga ttcctggggg aaatttttt gttttgctct tgtacctcat 72660
gtctggatta ttttggattg ctttggggac agtatctgag tttctatctc ttggcctgtt 72720
ttttccagga atataaaggt tttttttctt tgacatatgc ttaaatgttt atttttaagt 72780
gatgtaactt ttcaaaaaac ttattacagt ttatttctgt gggaaaaata ttttttakgt 72840
ttttgactgt tttttgttcc ttcttgtttg aaatctctag ccaacaagaa cattagtcat 72900
gacaagcatg ccatctgagt aagtacttgt tttgatttct gttcaatgta aaatgttaac 72960
cttttctctc ttatactcta attctgggtg cctttaggca acttgtcaat ctgtcctgta 73020
tcacttttac tttataaaat taatatctga gttagaagat cactgaaaat taaacatgta 73080
ccaaatgtga gcgacttagc cttgaaaact ctggggttgt ttaggcagca ttaagaggtg 73140
tgtgctcgtt ttggtgttct tttgcttgct tgataccaaa tagcttcatg aatgttcaag 73200
aagtggaaca tcattgacca aaacatttcc cttaaaggtc ttaaagcaat actgcagcag 73260
aaagctttcc acagcagtgt taaagttgct atgtatgcat tttgtggaag ggtcaatagc 73320
ttgttggcat gctcttatca tctcccttaa acatttaaca caacaaagaa catccaacaa 73380
aaatacagtg ctatattctt tgcaacagat ttttgaattc ctgtttaaag gggaaaacca 73440
tgtttttgat atcaatcata ggtttttaagg ttttaagaca tccatcaaaa cattggaaca 73500
tttcagtgaa aaatatgctg cagagagggc acctttagaa cattttcagt agtgggatcc 73560
ttttcctgcc tggggcttag aaataaaagc actgatcatc aaacaccata cattatatag 73620
tgaaaaaggg ggtcactcaa aatttttgta aatatattat gaaatatatt gaacattcta 73680
aatagtctaa tacagaagcg aatattgaat atatgtgtaa tatttttaa agtctttgta 73740
tttttccaaa ataaaagaaa aattactagt taactgctta ttttctcatt caagatttaa 73800
aaataaaact tttcatttag gccatcttct tgtcttactc tttttttctc cacatggact 73860
tcttgtgata cttaagaata agacctggac attctgattt tatgtggatt agctgagcct 73920
tgcagagaca cttgttactt actggcacat ccagcaagca gctgccagcc tcaggatgga 73980
gttctaggga gtgtgtagtt tagagctttt tacttttttgt ttttgttttt gttttctttt 74040
atcatttttg cctttatttc tttccaagtt taattatttt tcttgactca agcacacatt 74100
ctcgggttga agtagtgatg aggcccagat cttgactcac acatcttttc taccctaagg 74160
atctcttaag aatttaaaag catgatataa ttcagccctt tcatttttaca gataaagaaa 74220
caggttttga gatggacata cctaagatca ctagagataa aactaagaag gctgggtgtg 74280
ttggttcacg gctataatcc cagcactttg agggtcccag gtggacatat tgtttgagcc 74340
taggagttca agaccagcct gggcaacata gcaaaacctt gtgtctacaa aaaaatgcaa 74400
aagttagcca gacttggtgg tgaattgcct atagtcccaa ctacttggga ggataaggca 74460
ggaggatcac ttgagccctg gagatcaagg atgcagtgag ccatgattgt accactgcac 74520
tccagcctgg gcaacagagt gagaccctgt ctcaaaacaa taaaataaaa ctaaggaaca 74580
ccatcatttg gaaggaagag tgttagaggc agtctgtata agcatagaca ataacctctt 74640
ccccttttgta atataatttt tggagaggag agatgtttat ttctttttct atttatttat 74700
ttatttattt atttatttat ttattttgag acagagtctc cctctgtcac ccaggctgga 74760
gtgcagtggc gcaatctcct cccactgcaa gctccacctc ccgggttcac gccattctcc 74820
tgcgtcagcc tcctgagtat ctgggactac aggcacccgc gaccacgccc ggctaatttt 74880
tttgtttttt tagtagagac agcgtttcac catgttgttg tatatatcac agtgtggctt 74940
agaaagccct ccattgggga ttttttaaat tttctgggag agagggaaaa ctaatgtcag 75000
aactaatggc atagaaaggt tattataaaa gggaagaaag aactgagggt tgtttggtaa 75060
ggaagttgga cggaaagaat atatttttttt aaaggatatt ttaagtatta agggaatgac 75120
agagcaggag ataagccata atggtcatga gctttgtgac aaataggtcc cagatttgat 75180
ttgatgattt aataaaaagg gtctttttttc ccctcttagt agaaaaacta tgtgttgata 75240
ctcaataaat attacatttt caaaataaaa taagtgaggt tcttggttct gagcatgcac 75300
agataggttc aaataggcct gaaaaacaaa tcattgcccc agtgggaaga gtgttggtct 75360
gatgtcaggg gcctggttcc tttttttctt ttttcttttt ttctttttttt ttttttgaga 75420
cggagtctct ccctgtcgcc caggctggag tgcagtgaca cgatcgcggc tcactgcaac 75480
ctccacctcc cggattcaag ctattctgtc tgcctcagcc tcctgagtag ctggaacaac 75540
aggcgcgtgc caccacgcct ggctaatttt tgtagttttt agtagagacg gggtttcacc 75600
atgttggcta ggctgatctt gaactcctgg tgatccaccg gcctcggcct cccaaagtgc 75660

cgggattaca ggtgtgagcc accgcgccca gccaggggcc tggtttctga tgctggctct 75720
gtccctaccc agcccagcca ctgtgggaag ccattgacag cctgtgggct tgtcttctca 75780
gccattaaaa tagaattgag atctgaagtt tatttcccca ggtttcaaag cattgattat 75840
aagtcagtta agatatacgt accataacca aaatcagttt caaattttgg ctttctagtt 75900
ttattagtac taatattgag tgtaactgct ttgatgggca tgtgcaacaa agtcattcat 75960
tttgttaatt tttcccccga tttgacagaa agcagaatgt cgtcatccag gttgtggata 76020
aattgaaagg cttttcaatt gcaccagacg tctgtgagam cacgactcac gtgctttccg 76080
ggaagccact tcgcaccctg aatgtgctgc tgggaattgc gcgtggctgc tgggttctct 76140
cttatgattg ggtaagccct gtgtgtgaat gcgtatttta aaacaaggca ttttgataga 76200
gtgggtcacc ctgaggtgcc gacatcagca ctcaggccgg cgtgcaccct tgtggatctg 76260
cacactttcc tgtgagctgg gaacacccgt ctttcctcct gttggtctcc cgtgggctgc 76320
taccttcaa ccagggccaa gttctggggc aacaggagga cggggagggt agagagcagg 76380
aagtgagtag cctctaagat aaagcagaag caagattaca aagatgctga aagaaacgca 76440
aaatgcatgt tctcacagtc aaagagcttt cctctatgtg tgaccaagaa acattgtgag 76500
ctgtggtggt ggtggtttgc agagccaaaa taattcagtg attgtttgta cagatggatt 76560
tacttaggat gaaggatgtt cttttaatcc catttggata ggttttatcc tatgtatatc 76620
tatctgtaac attatttgcc cttgtttctg tagattaaag atagctttta aaaatacata 76680
attattttcc ttattcataa aaactgaaat gaactgttat tggttctatt attactttca 76740
tcctcaacct aaggttgctc caaagcattc ctttctggtg acagtagcat cacttgttac 76800
gtatgttacc attctgcatc tgtgggatcc gtcttccctc ctcctctccc aagaatgtat 76860
tctattcata ctcatactgt gttcatttaa accagtagaa ttataacatg caaaagctac 76920
acatgtattt tcaagaatgg ccgtcgtctt ttttccgtgt tgtgacagag gttaaagaga 76980
ttagtgcttc tagttgtgaa gtggaaaacg ttgaaattcc aaaagtaagc actgttcatt 77040
tgcattggtg gcaatggggg atcaccttac ctgattatat attagtactg ctttatgttt 77100
atttggatga aagacagtag tgcccctctc atccagggtt ttgttttgtg tagtttcagg 77160
taccatggtc tgaaaatatt aaatgggaaa tcccagaaaa taacaattta taagtcttta 77220
aatgcattct tttctgacta gcatgaagaa atctcaggtt atctggctcc attctccctg 77280
ggatgtgaat cgtccttcag tccagcctgt gcatggagta ggtgctgctt gccctcactt 77340
agtagccatc ttggttatca gatagaatct cgtgattttg cagtgtttgt cttcaaggaa 77400
cccttatttg gcctaataat gttccccaag cacaagagta ttgatgctga caactttgat 77460
atgccaaaga ggagctccaa ggtgctttct ttaagtgaaa aggtgaacgt tgtccactta 77520
atatggaaag aaaaatggta tgctgacgta gctaaaatct atggaaaaaa tgactctttg 77580
acctgtgaaa ttgtgaagaa ggagaaaaaa ctgtgcatac tatatatata gggttcagaa 77640
ctatccacag ttttaggcat cccccagggg gccacggact gtgccccctt tggatagggt 77700
ggactactgt ctctttaata actctagcat cagtgaatga gttctgtgtt ttatttctct 77760
ccaattcaaa tcgtctctgt gtcttcatct gactactctc ccttccctca ggttttggag 77820
gaaaaaatgt tatttctaag gatatgcatc tgtacaggat tccttaccca acttattctt 77880
ctgggacttg gagcagtcca tagaggtcag acgtgagaac gtactgcctt tgctgtcgac 77940
atggatagag acctgctccc tggttgtctg catgtctctg ctcagtgttc tgctagtact 78000
ccacagctaa tcatacatag aaacagaact gggtgaaatt ttaggttatt gtatctcttc 78060
tgggattacc tgatatgata aaggtgggca ttaaaacaca ttatttaata aacttctcac 78120
ctttagtcta gactccttgc ctggagggaa gaacctgggg cactcagaca cataagtgaa 78180
tgaatgaggt acaaggcaat cagacaagaa aagataataa aaggcatgta ggttagaaag 78240
gaagaaatag agttatctct atttataaac cacacaattt tctatgtaga caagtcacaa 78300
gcaatctaca aaacagcaat tagaggtgac agctgagttg agcaagtcat ccagatgcaa 78360
gaattccatt gaaacttcag tataaagcta ataaaataag tgcaggatct gtgtgctgaa 78420
aactacaaaa tactgatttt aaagctcaaa gaactaaata tattaaaaga catacaatgt 78480
tcatggatta gaagacatag tacagtgaac atgtcacttc ttcccaaaat gatgtataga 78540
tttaacacat tctcattcaa aatctcagtg gactctttca agatacagac aaactggttc 78600
taaaatttct atggagatat taaggagcca gaatagccaa aacaatttag aaaggaaaga 78660
acaaggagga ggactggcac tacctgcttt tggggcatcc tttcaagctg tggtcctcaa 78720
ggcagtgtgg tattggtgga cacacagaac agacagagaa tccagaaata gacccccaaa 78780
atacatccca tgggttttca caaaggcatg aaggcaattc agtggagaaa ttcagtcttt 78840
tgaacaagtg gtgctggagc agttggacat acacaatcaa gaaaaggaac cttcccaaca 78900
ctttgggtgg atcacctgag gtcaggaatt ggagatcagc ctggccaaca tggtgaaacc 78960
ccgtctctac caaaaataaa aaaactagct cggcatggtg gcacctgcct gtaatcccag 79020
ctactcagga ggctgaggca caagaatcac ttaaaccggt gagatggagg ttgcaaagag 79080
ccaataccat gccactgcac tgcagcctgg gtgacagaga gacaccctgt caaagaaaag 79140
aaaagaaaag gagaggagag gaggaaggaa gggagaacct cattctatac cttacacgag 79200
ccacaaaaat tacctccaaa tggatcatag acaaaattta aaggtataaa acttctataa 79260
gtaaacatac aagaaaaatg atcttggtgt aggcaaagag ttcttagata caccaaaagc 79320
atgatgaata acagaaaaca tagataagtt agatttcatc aaaattgaaa gcttttactc 79380
tgtgaaagat attatgaaga gatcagaaga aaacgtttgc aaatcttata tctgacaaaa 79440

gatttatgtc tggaatatat aaagaactct taatactgaa caataagaaa acagaacagc 79500
tcaaacaaaa aatggcaaag aaaagatttg aatagacagt ttactgagga cacacagatg 79560
gcaaataagc atctaaaaag atgctcatca ttattgctca cttcagaaat atagtgagat 79620
ccactacata tccattagaa tggctaaaag aaaaaataac agtcgcactc tagcaaggag 79680
ccagggcagc tggaacggct gctggtgcgt gtgggaagtg gtccagccgc tttgagaaac 79740
agtttgacag tttcacagaa agctaaatgt ccactcagca gtcccactcc cagatatttg 79800
cctcggagaa atgaaagctt gtgttcacac agagtctgta cgcgaatatt tgtagcagcc 79860
ttacttatca tcagctggac ctggaaacag cacagctgtc cctccagtgg gtgaatggat 79920
caaccagctg gaccaaccat actgtggagt gtcactcagg agtcgaaagg aatggtgata 79980
ggtacagcag cttgcatgac tctcaggggc atcatgccaa gttgaatagc tggtctcaga 80040
aggtcacatg ctgtataagg ccatttcttt gtcattctag acaaggccaa actataggga 80100
aggagaacag atgagtggtt gccgcgcatt aaggtgggag tagcatctgc ctctgcagaa 80160
caatagcagc tgtcacatct ttggggcatt ggaattgtgc tgtgttgtta gtggcaatgg 80220
ttacagaatc catgtattaa aacacagaga actgtacaca catatgcaca cacgagtaaa 80280
tcttattgtt tctaaattta aattaaaaag aatatctagg cggggtgcag tggctcatgc 80340
ctgtaatccc agcacttttg gaggccgagg cgtgtggatc acgaggtcag cagttcaaga 80400
ccagcctggc caagatggtg aaactccgtc tctactaaaa atagaaaaat tagctgggca 80460
cggtggcagg tgcctataat cccagctact caggaggctg aggcaggaga atcgcttgaa 80520
cttggaggga ggaggttgca gtgagccgag atcacgccac tgcactccag cctgggtgac 80580
agagtgagac tctgtctcaa aaaaaaaaaa gtatatctta catatctaac gtgctttcca 80640
aatggagatg tttgagcact ggtaggaccg ggctagtgtc ttggtttcag aactaggttt 80700
ccttctgtgt gctgaagttt acaggctcct gtaccttcaa ctgctgcctc tgtacctata 80760
cttcctgtta gcactgaagc ttcatcccag cttttctatc ttaaaaaaaa aaatgaaaag 80820
aatttaaaaa cataactttc tctaaattgc tctttgccct ctgtgctacc ttttttttccc 80880
ctcattcatg gcaaaacgtc acaaatgtat gtctgtattg cccttgcctt actgatgatg 80940
tcgctatttg ttaatagtat caactcttgg gagattgcga aggctcaggt ggcctatggc 81000
ttcaggtgaa atatctgttt gtgtgattac aaggtaacca tgatggcagt caggtatatc 81060
acacatatat aaatgacaca aacagatata aatatatgtt tgtgtgatta caaggtaaac 81120
gcaatggtaa ccgcaatggt aaccacgatg actctcgctg gcacaacagg agtattgatg 81180
ttcacaggtt gctcctgact tgcaccctca aaaagtttag aaacaagccg agtcactttc 81240
tctgttcatc tcmgtcttca agaagacaaa gacgactgct gcttcttgca tggccccccc 81300
cctttaactt ttaaataaat tgaatagtac aaacataaga aatttgagag aggatagttg 81360
ccaccaccat ttacaaagcc attctacata atttttaaag cttagcaccc actttaatat 81420
ttatctatgt cttgcatata acttcagata taaacttcac agttccaatt tcttttaggg 81480
tcaagattta aagtatccat atcatatatt atatacattg actttgtgta caaggaatct 81540
ctctctctct ctctctctct ctctctctct ctggcactct cgctctctcg ctctctcgtc 81600
ctcctccttc taaccctgtc tccaatgtag ttgggggatt cttaaaatat tctctttggc 81660
tagcagtata aactggcctc caagaaaaac actgctgagc atgtttttat ttcagggttt 81720
gtgtggtatt ctctggaaat ttcttgtaaa ggagatttgt agcagttctt cagaattaga 81780
tggttgtatg tggcccagct agtcttatca gaaactgtgg cgattttata acaaagttca 81840
gtttgaattt tgacttaata tttttgagaa gtttattggc aatttttcca tgtttacagc 81900
agttcacacc tccagtgtta gcgctactgt tttcaggaaa gagaataatt tatgtttttc 81960
ctccttcatg actgaattgt ctggcagata catggaaata gaaaaccatg ccaggagttg 82020
ccgagcttcc tatttatggg agacaggaag taacacaaca gaaaaataaa gaaattaatt 82080
tgaccaaagt gtccctttag actcacattg ttttgttatg tgttgttcaa gcatagcaca 82140
atttgaacct ttaaatactc tttatcccac tctcacttaa tttgatgttt cctgcacttt 82200
cctgtgactt gtctaaaatt ctactttccc tcgaaaccct tttgtggatg ctaacataca 82260
agcagagtgt cctgtgattc agtcttccct ttttccagct accactccgt gtcactctgt 82320
ccagcacagt gaggaataac tcagcctgta ttcagatttt aatattttga ttctgaacag 82380
cttatgaaaa ggatctgata atagagattt aaagctaatt cacttataaa tacaagtgta 82440
gggcttaaaa gctaaatcag ctttacaaca aaatgtcaag gccgctaact atcaacagat 82500
aatctagtgt tttcttaatc aaaaatgatg tcatgatgac tattttcttg agataatgtg 82560
atccacattg aacttagtaa gcagtgagtc agatgagata tgtttttatc agtggtgagc 82620
atagaatcaa tgaactgtta gaataacaca ctcagttcat tccgttcacg cgtctcattt 82680
tacattaaag aaatgctgag ccgctctcct aaaattataa ctcatggcag aaccagaact 82740
ggaatctcag cttttcactg gtgttagttc atcaccctgc attcctaagt ctgttcaaaa 82800
gggatcatct tgaaaaacca ttctcttttt aaccttcagt tggcagatta acttcataac 82860
tcatgttagg aagaatcttc aggcacattg tacttggtgt gtcacactga cactgagttt 82920
ctgagggtgc ccttcaggtc tctctggcag acatttattg ctcgcacttg caagctgact 82980
aggatctcag gcctgggtct ctgaactttc acggcttgat ttcaaagtcc tttttatcct 83040
gctacagatt ataccttggt aaaggacttt atacttcaca gagtgttttc acatgcactg 83100
tctcactgga tcctgacaga acatttttgc agccgagaag gacgctgcaa ataattagtg 83160
agtttagtga tggagactct gggcaaaaat agcttgtctg acttgaatgt ggatcttaga 83220

aacacatctc tgtcaaggca ttgttttaag gcagtgacta tggtcttaca tttatctcca 83280
ggacacctaa tttatacttt ttcctgatta aaataatgga ttctggtttt gcccagacat 83340
agaacccaca gagtttgtct gcttctttca cttgaggtgg ttcctgagca gtgccagagc 83400
tcattctctg cggaggctcc tgcaggctgc ggcagcgtgg cctctggccg ctgggagcat 83460
gggaagcagg cgctgcggtc taggtcctcc atccccctgt ctgctgctcc tggcaagacc 83520
ccaaggtgcg catttcccag gttggagccg ctgtgcttcc caggaccata atctgctgat 83580
tgaggacaga taccaaaaag tgattcatct gtaaaattga gggctgtggt gctgccctct 83640
aggaggacat ttggaaagat gtggagaaac ctgtgagtgc taagaatgac tgatgttaaa 83700
gtttgaaaga gtcaaagtga ttttttttagt gggagaagac tgtggagtca ccctgagatg 83760
caaccacagg cttgattaga aataaagttt gatcaccatt ttcaaatttt tacattaata 83820
ttttttaatt ttcgaaaggt gctaaacaga atctacttaa tgcacctggc acagaaaagg 83880
cagtgcccgg gtcctaaggc tgcacctttg caagaaagag maatacctga ggcaccggga 83940
gtgaggagga caggtgttgg agaaggctgt agggccccag tatggctgtg tagttcaaga 84000
cgagggatgc agaagccatc ggactatttt aattacagag tggcagcttt tgtctctgtg 84060
gcctctcagc aaagaatgga ttgcagggag gtaagaacag ggtgagaagc aggaggcagc 84120
tagggtcatc gaggtgaaaa atgactgcgg ctgtgtctag agggggggtt gataggtgga 84180
gaggagagag caggtcggcg cccttcctag gaagatctag tggaatctgt aacgtcaggt 84240
gtgtgggaat ggagaagtca agaagactcc cacccaaatt ttttcctggg gcgactaact 84300
atagataatg gtgccatttg cagagttagg gaattctggg gcagaagatt gtgtgcaagg 84360
tttggggtac aataaaaaat tgatgtaggc atattaggtc tgagattcct actggacatt 84420
caaatagaga tactacatat cagattatat atatgtacat atattcagag gaaaggttaa 84480
ctattcactc cagccatggt acctggaagg gagtgtgaat gaagaaatga agaaaacagt 84540
gagtttaggt ttgatctctg ggctgtgccc tatgcagaag tcaggggaa gggggaggca 84600
gggggacccg ggaacggcta gctagcaacc tgggggagac accaggggaa catggcatca 84660
gtcagaaggg ggactgtctc aggaaggaag gatgctcagc tgtgctgagt gctgctggaa 84720
ggtgaataag aggagacaga agccactgtt tgatttcttc aggtggatgt tgtcagagac 84780
cttgaaaaaa gcaggatgaa tccaatgact aagacagttg aagagtcaat ggtacataaa 84840
gcagtggaag cactagggtt atgtgtaatg gtgcgatttg ctgagttagg gattattatc 84900
agacatattg ctgatatgtt attcctagac ataatgctgc tgctacatca gagagattgg 84960
ttggcagcga atggggcact gtgaagtgtg actcgagcct tctcgtgttg ccaactgcaa 85020
cacagatcat cgtcctagtg cttggcgatg tggttgcatt atggtgagtt gagtgtggcc 85080
ttgggaagca tctgaatctg ttggctgagt tatcagggaa aaaaaattta aaaagtaaac 85140
taagattatg tatattaatg aaaaagttgc tgtatttggc aaatacttta aatggataag 85200
gctaaaaacc aacaagtcga gagggtactt gttgccaccc atccttttcc aaatcatggc 85260
cttcaaggat cacactgttg gtctttcctt ttcttttaac ttggatcaac tgtgaagtaa 85320
cacaggtctt cagtgtagat ctcagttccc caacatttgc cttatgactg agacctccag 85380
gacgtcaact tggtccatgc tgaactgcag cacaaattcc aagctttgac catacctcaa 85440
ggtgcacttt aacctttgca gtgttctgcc agacatctga actttcactt ttgtttctga 85500
catctcaatc acacagttct cactgtaaat attaaataat agcacagaat attttaactt 85560
caggtattca ttggaaaatt caaccatggt ttggttttat ctgtcacttc aaaaactgtc 85620
ttcagctgtc catcatttag atgtcattta gatgttcctc agggactttg gggacattgt 85680
taacaatctg ttatttcaag gcttctaaac tctatcccca agttaaaatg atttccaagg 85740
aacatcatac ttctcttaca gtctgtgtgt aagcaccctc tgtgaattcg gttttaggga 85800
caatgttagc ttttgaagag agctgatgta agaaatacta gattttagga aactgttgta 85860
ctttttttcaa agctatattt gacgacattg tacattttgc tacctgatac ttttgatgta 85920
tgatccacct aatgcctttc tcctaaaatt aatttccagt gaattgaata ggaattccaa 85980
atgaaatgaa tttcatagga aaatctcata cagaaaattt gttaggctgt ccttaaccag 86040
agaatgagaa ttatgtaatg cggttttgtc agctagagta acagcttgcc ataggttcat 86100
aatagagctg ttttttagtt ctttttcttg ggttcttgtt tctgaaagaa agtttctctg 86160
ccagaatatt gaagtcgtgc ctaagttaat aatttaacaa gcattgtata tattaataat 86220
ataatatcaa taattaatgc tattaatcat taataacaat tatttaatat taatattaaa 86280
tacttaatat taaattttta gaatattaaa atttaaaatt taaaaaataa aatttatcaa 86340
aaaaaatttt ttttttactt ttgaagcatt ggttttatta aactttcaaa gtagtatggc 86400
aaaaaggtgg ccacatacca aatagtgtca tacatttctt aaaatctctc ctagcaaata 86460
aacttaaatt gagatcatga gtcagttgaa aagacaattt aatttttttg ccatacaatt 86520
aaagtatttc tgagaagtca gagtgctttg caatgtttgg tgaataattt acacaattcc 86580
agaataatgt ctcacttatg gagaatacac ctaccactta cttcgataaa cagaagtaga 86640
gtctatggtt tctttctttt ttttttttt ttttagctgc taaagattat tattaggaca 86700
gaaggacaat tagctttaaa agcattcctc agaacatgta ttttttttc tagtattctt 86760
tttttttat tatactttaa gttctagggt acatgtgcac aacatgcagg tttgttacat 86820
atgtatgcat gtgccatgct ggtgtgctgc actcattaac tcgtcattta gcattaggtg 86880
tatctcctaa tgctatccct ccccactccc ccgaccccac aacaggccct ggtgtgtgat 86940
gttccccttc ctgtgtccat gtcttctcat tgttcaattc ccacctatga gtgagaacat 87000

gcggtgtttg gtttttttgt ccttgtgata gtttgctgag aatgatggtt tccagcttca 87060
tccatgtccc tacaaaggac atgaactcat catattttat ggctgcatag tagtccatgg 87120
tgtatatgtg ccacattttc ttaatccagt ctatcattgt tggacatttg ggttggttcc 87180
aagtctttgc tattgtaaat agtgccacag taaacacacg tgtgcatgtg tctttatagc 87240
agcatgattt atagtccttt tgggtatata cccagtaatg ggatggctgg atcaaatggt 87300
atttctagtt ctagatcctg aggaatcgcc acactgactt ccacaatggt tgaactagtt 87360
tacagtccca ctagcaatgt aaaagtgttc ctatttctcc acatcctctc cagcacctgt 87420
tgtttcctga ctttttaatg atcgccattc taactggtgt gtgatggtat ctcattgtgg 87480
ttttaatttg catttctctg atggccagtg atgatgtgca tgttttcatg tgtctgttgg 87540
ctgcataaat gtcttctttt gagaagtgtc tgttcatatc cttcgcccac ttgttgatgg 87600
ggttgttttt ttcttgtaaa tttgttagag ttctttgtag attctggata ttagcccttt 87660
gtcagatgag tagattgcaa aaatttttctc ccattctgta ggttgcctgt tcactctgat 87720
ggtagtttct tttgctgtgc agaagctctt tagtttaatt agatcccatt tatcaatttt 87780
ggcttttgtt gccattgcat ttggtgtttt agacatgaag tccttgtcca tgcctgtgtc 87840
ctgaatatta ttgccaaggt tttctatgct atagaaatag catatttcta tgctattcat 87900
cattaataac aattatttaa taatattaat attaaatagt taatattaaa tttttagaat 87960
attaaaattt aaaatttttt taaaaataaa tattttatat taaattatca aataaatatt 88020
aataataatt atttaatatt ataaaattaa taatctttca ttattgaatt attgattgag 88080
ttaagtaatt aattgattaa ctgataagga ttattgttaa attattgtac tcttgggtag 88140
tacagagact gcatactgcg ctttgccatg taaatactat tgtctacttc ctggtacgtg 88200
gctctaggga ggctatggca gagtcaagtg cttttgccct taatgtgaac aaaaaatagt 88260
gattgctctt agtagccata atatttggtt tattgtctgt gttggtaata atttctgctg 88320
tgttttcata cagtgaagtg atgtttctgc tgtttatttt agttgcattg gaatttgtta 88380
tatttatttc tttgttttcc ttttgataag agaagtacgc acttagttat ttataaagat 88440
gtttggactt cacatgtgag tacagtggtg acatgctggg ttttcctggt cattgcttag 88500
ctgtatttat aaagtgaata ttactgagca gttaagcctt aacatcgaga atcacccatt 88560
ttcatttttg aaaactggaa aggattaggt agaatgcaag gagaataaat tgaacttaaa 88620
tgtttgtgtt caattgaggt gagctttttc ataagaatat tcaagcctag gtcaacatgc 88680
agcttgtttt ccctctcacc acctggaatt cagtctctat cggtcaatgt cttctaaaag 88740
ggaaatgggt tcttaactat atacttttag tactttattg cttatcttcc ctttcttggt 88800
tgaataggct gtgttggata tttagcttcc tgccccttte tttatgagac agctagggca 88860
gtgcttttca aaaccttact aatgtgtgga tcacctgggg gatcttactg aagtgcagat 88920
cctggttcag tgggtctggg tctgctcagg cttgaggtga ggtccacgct gctagtcctg 88980
tgacccagca ttaggtcccc aggatacaaa atatgaccgg ggatctctgt cgtattcggg 89040
ggtggagatg agacagcgtc ccaatgatgt tagtcacatg gaacatttag agatgcggag 89100
tactttgtca gtgttttaca catcgtcaag ctgttagtca agacagtaat cctctgtgga 89160
aactgtgggt tgaacacttt cagtaaattg ctcatggtca tagtgcttgg aaatagtaaa 89220
tttttttttt tttctttgag acagagtttc gctctgttgc ccaggctgga gtgcagtggc 89280
acgatcttgg ctcactgcaa catctgtctc ccaggctcaa gcaattcttg tgcctcagcc 89340
tcttgagtag ctgggattac aggtgcatgc caccacacct ggctaatttt tatttttttgt 89400
agagacagag tttcaccgtg ttgtccaggc tggtctcaaa ctcctgacct caagtgatcc 89460
gccgaccttg gcctcccgag gaactgggat tacagatgtg agccactgca tcctgccaga 89520
aatggtgaat tttgaatttg aattcagctc ttcctcaatt catagcccac attctttcta 89580
gcatctactt ccaaagatag cctagagagt atttttttatc ttctatagct gtaaaccttg 89640
atatgggcat tctctgatgg cctgtgtgtt ttgaaaagat taatggataa ggcagtggat 89700
ttcactgcta accttgctac accgtagctg tgtaaccttg ggtaaggcag tttctttatc 89760
tgtaaaagaa tggaaagatc acctaaataa agtactcagt aaacactcaa taaatattaa 89820
atatcgttat tattcaacaa gcatttttga cgctgatcac tagccttcat taaaagtata 89880
acttggatga acgttgaaca caccgagtga aaggagccag acacaaaaag cacatgttgt 89940
ataattcctt tcagacagta tatccagaat aggtaaatcc atagaataga aaactaatta 90000
gaagttacca gggatggagg ggagagaggg atggggagtg attacttaac aggtacagga 90060
tgtttttctg gggtgatgaa agcattttga aactagaaag aggagctggt tgcaccgcat 90120
catgatataa aatgccattg aattgcacac tttaaaatgg ttaattgtat attatgccaa 90180
tttcacctca cttaaaaaaa gtcatatatg gaaaatagct ttaaggcacc actacaacta 90240
ctaaataggt ttgtattttt aaaagaactt tatggaatta taggaagcat ttcttgatgt 90300
tatgagatgt gttggaaata cagaagaata gcttattttg gaacagatat tattggcttg 90360
aaattttgcc agttcaagct ggtctctttg gaagactaga cctttatttt ctggcttgaa 90420
aatgctttgg acataagtac cctattattt tgttgttaaa aattatacta ttgacatccc 90480
caatttttte tcctgaagtt cagtataacc tagaaataac ttcattgcta cactatttca 90540
ttaactacat gggtgctttt ttagttaata atgatgcata atgtcttcat gtggcagaaa 90600
cactaacctg ccccttgtca taaatctgta aaaagatgga cattggttta aacccagttg 90660
ttgaattctg tgcctttaac cagtatgtta cactgtctag ttggggaaga atcccaaatc 90720
ttcttctttc tttagaaaaa tccaaaacag catacaaact agcaaactct cataaatgtt 90780

gtttgagaaa atcaattgcc ctaactacta agacaaagga tctataaaat ctgatgagaa 90840
caatctttgt aatttgattt ttataatttt gtcagcttaa attagtaaaa agttaataat 90900
tattactttt gttacgctta taataaataa tgtgtttcta caccttccat aaacacctac 90960
aaccacactt tttaccacag ttggtggagt gaagggtgga tggaggagat agtggcaaaa 91020
acaccccaat cactttcagt gattaaagta aagatgtgtc taactttact cctaaagtat 91080
catccagtaa agtggaatgt aaaacatact tttgaactgt ttgaaatcaa ctacattcct 91140
atggcttacg actgtgggac aagtttctaa ctatcagatt tgatttttaa ttaatcagtg 91200
atattttata ccagcagtct ccaacctttt tggcaccaag gaccagtttt gtgaaagaca 91260
atttttccag ggacttgggg gttgggaggg tgagggagga atggttttgg gatgattcaa 91320
gcacattaca cttattgtac actttatttc tattattatt acattgtaat atataatgaa 91380
gtaattatac aactcactgt aatgtaaaat cagtgggagc cctgagcttg ttttctgcaa 91440
ctagacagtc ccatcggggg gtgacgagag acagtgacag atcatcaggc gttagattct 91500
cataaggagc atgcaaccta gatccctcat ctgcacagtt cacaataggg ttcgcacttc 91560
tatgagaatg taatgccact gctgatctga caggaggtgg agctcaggtg gtaattaaag 91620
caatgggaag tggctgtaaa tacagatgaa gcttccttca ctggttcacc caccactcac 91680
ctcctgctgt gtggccccgt tcctaatagg ccacagactg gtaccaggac ccctgtttta 91740
cacgatgtgg agtcttttgt atgcaaagaa tattgttgac tttcgccaca cggaagcccc 91800
cccgccccgc ttcccccgcc tttttccttt ccagttacat tcccacaggt attcttagta 91860
ccacaactgc agttgaattt cacagtatgg tgggtggtaa gctatggtgg gcggtaygct 91920
tggataagcc tggctattta gaaatttgga ataaatgtag tgttatgact aacagtaatg 91980
ttgcctatca aaaattgtga atgttaataa atgttttcaa cacaatcatt aatgctttcc 92040
agtgagttaa accagcttca tgttacagtt gtattttcca tcccagtagg gagtcattat 92100
taaatggggt catgttttca agcccaactt aaaatccctc ttacagattg ccttccccac 92160
cccaccccca gttttctctc atcacttata cattgaaata attgcttatt gttttccctc 92220
tttaaatttt ttttgagaag tcaaaaattg agtaccttgt tcagtgtttt tgcttatgaa 92280
atactttgtg aataaatttt gttcttagct gaagaaaatt tcttaggcag ttaagaaaat 92340
actaataagc taattaatga ataaaaacta atttcattgg tcctgattgg aagtgcaaca 92400
tttaccgata tttagctata atccttttga tcagtcagaa atttgtaatt attctttgag 92460
aaataaaaag ttgagagggc tgggtgcggt ggctcacacc tataatccca acacgttgag 92520
aggccgaagc aggtggatca cttgaggtca tgagttcgtg accagcctga ccaacactgt 92580
gaaaccccat tctctaccaa aaaaaaacaa aaaaaaagaa aaaagaaaaa aattaaccag 92640
gcattgtgat gtgcgcctgt agtctcagct acacaggagg ctgagtcagg agaatcactt 92700
gaacctggga gacgatgctg cagtgagcca agattacacc actgtactcc agcctgggcg 92760
acagaggaag actgtctaaa aaaaatagaa aaggaagttg aaaacagctt agggaagagc 92820
tgcaaccact gaccagcacc agtactccat cataatatat gcttttcact tataaggaac 92880
tgtaatgtaa actgtggact ttgggtgata atgatgtgtg aacacgggat gactgggtac 92940
aacacatgta gcactccagt gggagacatc aaaatgcata tgtggcggca ggaggtgtat 93000
gggagctctc tgtaccttcc tcttaatttt gctatgaagc taaagtggct ttaaaaatac 93060
aaatacagaa aaaaacttgt gctttctata gattaatttg aacatagaca cattaatata 93120
atagatacat tgatttgaac ataggtacat taagttgaac acttaaggtt tttatgatgt 93180
cctataccac aataaactga agaagtctgc cttacaaatt tgttcaaaga actctcaatg 93240
ctctcactgc tccttccctg ccttgaacag gaagtgtcat ccagtgcaat aaggggggaaa 93300
ataaaatgtg catagcaatc agaaaggaag aaataaagca gtttctattc acagatgcag 93360
ttcctattta aattcatcag caaggttttg gttttatgaa tgataatatt aaaatgtaaa 93420
aaacactatt ttcattatgt aatgtgtcac ctacaagatg ctgaattcct gttgcagcgg 93480
atgctgaatt cactctgccc ttcttataag aaatatgttg ggccaacctt ttgtttttaa 93540
gtttgcttac agccttacct gtgctctttc aaagtagatt ttcactattt tgaacactct 93600
attaaggtaa agatgtgttc ggccaatgaa actactagag caaaatgttt acactgtatt 93660
tctgatttga ttgttttaat acaactgaat tagtgttttc tcctatctct atgcaatatt 93720
aattcctggg atgtctgtgt aaattaatta atttactgac cagaactcta ctttagcttc 93780
ttatggtttt gttttcttaa catttagaaa cggctaaatt tagaggacat aaattttctc 93840
catgagattg tttaaattca gttgactttt taatgtggat tatatttgaa cttgaatgcc 93900
gcacgcattt ttaatgctgg ttcatggctt ctgtcactgg tacgttgtat ttctcactgt 93960
actattcttt tacgttgcct cttgtctgaa atgaacttga ttttaacctt ttatttttctg 94020
gtctaattat atgagcttgt ggggagcctc acatattgtt agtatatctc cttaaataac 94080
atgcattgag gctgaggtca gcagatcact tcaggccaga agttcgagac cagcctggcc 94140
aacacggtga aacccgatct ctactacaaa tacaaaaaaa attagccagg tgtggtggtg 94200
ggcgcctgtg gtcccagcta ctcaagaggc tgaggcagga gaattgcttg aacctgggag 94260
gtggaggttg cagtgagctg agattgcacc actgcactcc agcctggatg acagagtgag 94320
agtttgtctc aaaaaataaa taaattaaat aaataaataa aataaacatg aattgtataa 94380
tccagctttg ttattttagc tctaaacttc tggtgtatgg agacagattt tcagggagtt 94440
tggtcctgga ggagagacgg ctgcagaacc tcaaatatta ctgaattaaa aaggaaaaga 94500
ttgtattgat cattttaccg tgtggggatt caaatactaa gaggataatg atgatgataa 94560

tgatgacgat gaaagcttgt ttatgggaca ttttactctt ccaaagtctg ggaaggaatt 94620
tcaagtgtat tctggggact tctgaaaata ttagccaatg ttagaaacaa agtcgcaagc 94680
caaagggatt gcttttgaat ttaggcttgt gatccatctt cttttaattc actgttttaa 94740
ttaataaaag tctggaatat ttacagagga ttgtttataa aacttcacaa attagaaact 94800
tggaattaaa aatatatata taaaatattt catatgtgta aaaacaggat aatatttaaa 94860
tatctgacct catgagaata atgactcaga tttcttgtta tcgtgagact ttttctcaat 94920
caacttttta ttaatattca taacgtttat gcaacatgaa gattctgaag ggactttgtt 94980
gtctgagaac acatctattt cagatctgcg gagtgtatca ctttttgctg tgtcttcaaa 95040
gtgattcttg gtttattgcc tgctaaggct aataaatgta taataaatct gcttgttgtg 95100
tcacttgcag gtgctatggt ctttagaatt gggtcactgg atttctgagg agccgttcga 95160
actgtctcac cacttccctg cagctcccgt aagtcagatg ttgttttacg atggtaaatg 95220
cagtttgctg ttctcaagaa attattataa acataagggt ggacttaagt ttttatccag 95280
tcaagcacaa ttatgcccat aattaaaaag acattcacag aacttaacac cttttatcaa 95340
tttattcgyg agaacaaatg tgagaacgtg agaccactgt gcaaaaagta gtgaggaatg 95400
cagtccaaag aaaatttgac gattaacatc ctcagaactg agaaaaacaa aaatgaaaaa 95460
agactgaatt cttgggcagg tagtcttata tcttgcttaa tgtttttact kttaatagaa 95520
atagaactga taggtataaa gattatggct tgctggtgct gtgataacag tatttatatt 95580
tttatggctt tcctaaattc cacttcaact ttcaaatgct tcattgaaaa gttctgggtt 95640
ctaatttttt ttaagattaa gtaataatta agtggataat ttaaagtttg cttggataca 95700
ggattgtgca gaagttgcct ttcctgttca aaaatgttaa tttgtttgtc acagtttatt 95760
cattcaaaag attaatagct gaaagataaa tggtgatttt tatctgccac tggtgttgtt 95820
atttagctgt ttgagtaggc catatgacta aaacataaca aggagttgaa ctgtgctccc 95880
tgatcactgt agttatctag gttgttgggt tgttttgttt tcatttttaa gattactgtt 95940
tgatttcctt tcagctttat aaacattttc ttaaggagag acaaaagctc ctctcagcaa 96000
aactgtttgt ttgaaataCC gtgtaaggaa ctgaagtgta aagtaaaaac acaaattccc 96060
cccattctcg ctcataagag attatatatg atgcacaatg acataatgag atttgtcctt 96120
gaatttttta tcacctgcct acaaagagaa ttgatataaa ttgtgttgtt gccagttttt 96180
cctgcattar cgtttcccta cctaagtatc catcactctt gtcattgaga tatcctagaa 96240
acttgttgtt gtctttcgag gctgtgaaat tttcttattt tcagttgttt ttcaacttga 96300
tacaaggcca tgataccgtt gttgaattca taaaaccttc ttaaatataa agtagataca 96360
gttctaagat agggaggttc ttaactagtt aaatagttgt tggaaaagtg caccttggtg 96420
gaaataaaac agagccttga ctttgccaga gtccatcatt gactccaaat atgtagcaac 96480
acctgtgtgt tctaaaacta cgtcaagtgg tggggagaag ttggggtaaa ataaattaga 96540
ttttgaaatg gaataaagaa aaaataatgg tagaacactg taaggtgaag acagacatat 96600
agtagatgct agttacagac tggactctga acttccttgc aaatgattca gaaaagaata 96660
tatgagaaat tgcctttaaa ttataaagct ttacacaaat gttcattagt attaattgta 96720
ctatgaaaat ttcaaaagga gttaaaactc caggagttta tggttttgta gtcccgagta 96780
taaagctgtg ttctcaaatt ttcttttctt tctttttttt ttttttttt tccgagatgg 96840
agtttgcttg ttgccccggc tggggtgcag tggtgcgatt tggctcactg caaccttacc 96900
tccctggtgc aagcagttct ccctgcctca gcctcccgag tagctgggat tacaggtgcc 96960
cgccagcacg cctggctaat ttttgtatta tttagtagag acagggtttc accatgttgt 97020
ccaggctggt ttgaactcct gacctcaggt gatctgccca ccttgcctac caatgtactg 97080
ggattatagg tgtgagccac tgcgcccagc cctgtgttct caaatttttg gtaaatattt 97140
aaatatatta tgaacatcag attttgtttt tgcactttga aacccttttt ttttttcag 97200
tttgctgatt gacataaaaa aacttactag tgtcaattat ttttttcctt aagtaaattt 97260
aagggtgaat cttgagacat atagctttgt aaawttctta aatagaaggc ttttctcaac 97320
cagaaattaa attgtagtct agttctataa aaatatatct tactaggaaa gaaaacagac 97380
ctctgtttta gaatagtgag aagatagtaa agtttctttg tcatagaatg aaatgtataa 97440
ttttcctcat cattaaaagt aagaagtttc cttatcacaa ggcacaatta ggtcttttgg 97500
aaacaaatta taaaattgta aatattatca taaaagttaa acataggcat atcccctaat 97560
aagttatatt taattactaa aaatacCttc atatttaaca atcaggcaga aaaaaatagt 97620
acggtctgca tataaactaa aatggcacgt ttctgttgat aatttcagag attctggaag 97680
tttctaccat ataaatttga aatacgtatt tgagcattaa cttataacta agctgtcaac 97740
ataaatgtaa atacgctgtt tttgaaataa aaatttaaag cacctaagag atggagtaaa 97800
aatgcactaa ctgtttttcc aaatattaaa cttctagtaa ccccttctca gaatatccct 97860
gaatatgtct ttttatggct tagagagttt ttttcttcct tttaattgtg atagtgatgg 97920
tgaattcagg acatatgggt atttacacag tgtataaaca gtgctcagaa gaatgcagtt 97980
ccaagatgat ctgtattgta taacataagt gttctgtttt ccakttattt actgataaac 98040
ttgcacataa cattcttggt tgtgacagca gcgtctgtaa actgtcagtc tgattctcag 98100
cctcgggttc atctttgcat aggtgttctg tctaatcaca attatggatg tttagggtct 98160
tgctttggtc cgttaagtga tgcaagttta agtgataaag tttacaggct ctaatctgga 98220
gcatgtgggt cccgtcagca ccgagcacac gccctctgtg gtggaagagg acacagtgcg 98280
caccgtgact ttcagtgcac tgggcttaag tctttgaaaa tagttcgaga cagttcctca 98340

ggtggactgg gatgtttaga aatctgctgg tcggatcatc atggttgtgg ccttgagcga 98400
atagcctgag cctttccagt agtaccattt aatgccgttg aacttatttg tgttctgcct 98460
ctgtggatag tacattccgt tcaagttgga aggaccacat gcatcaaacc accagcctgt 98520
gaaagtaaaa cacagaagga attaggaact aggtgatgcc agctcccacc acgaagacag 98580
caatactcag ctaaggcagg aggcacactg caggcgtgtg gagtaggcac atgcagatga 98640
tggtgagtat aggatgtgca ctggcagagg gattgttttc cagccataca cccatgacat 98700
cacagttcca ttacggcaaa tgcttttaca agccttcttc cacctttcc cttgtgctgt 98760
gtggagaggc ctgaattctc cacagtccta tttggtaagc ccacagtgtg tacacactta 98820
cagcaggagt aagcaaacat ctgaggcaca gttggaaaac tctccttcaa ccaggattac 98880
tttgcagtcc cagcaacatg gtgggctgga ctcrctcagg ctccccttgc tctattaaat 98940
gatttttcg gttgaagttt aamctaaaat attaagtact cagtggagct acataaaaag 99000
gaagtctcta tgtttcagag acaaaaagga aatttaaagt gagagtgtgt gctcgctcag 99060
ctaaagccag ggcaggagag gtgtccagca caggggctgt gggagtgaag ccccatctgc 99120
accttaattt ctgggcttgg ccaaaaacag gagcatgctg gggtttgtga gagaaagaaa 99180
cacagtagtc ccccttatc tgctattttt gctttctgca ctttcagata cctgaagtca 99240
gctgggccaa aaatattaaa tggaaaaatc tagaaatatt ctataagcag gggccgggcg 99300
cagtggctca cgcctgtaat cccagcactt tgggaggccg aggtgggtgg atcacgaggt 99360
caggagaccg agaccatcct ggctaacacg gtgaaacccc gtctctacta aaaatacaaa 99420
aaattagctg ggcgtggtgg cggacgcctg tagtcccagc tactcgggag gctgaggcag 99480
gagaatggcg tgaacccggg aggcggagct tgcagtgagc cgagatcgcg ccactgcact 99540
ccagcctggg cgacagagtg agactccagc tcaaaaaaaa aaaaaaaaaa aaagaatgta 99600
taaaccttaa attgcatgcc gttctgagta acgggataaa atctcctgat gcccacttca 99660
tccctcccag aacatgaata atctcctcta tccagtggat ccacgctgtc tacatccggt 99720
ctcctgatca cttagtagct gtcttggtta ttagatcgat tgtcacagta tcgcagtgct 99780
tatgttcaag taacgcttat ttgacttaat aatggcccca aaagtgcaag agtgatgatg 99840
ctggcaattc agatatgtca aagagaagct gtaaagtgct tcctttaagt gaaaaggtga 99900
aagttcctga cttaataagg agagaaaaaa atcgtacact aaggatgcta agatctatag 99960
aagaacaaat cttatatcca tgaaattgga agcaatatgt tgtatattat tcagttttat 100020
tattgttaag tctcttgtga ctagtttaca aactaaactt tgtaagtatg tgtgaacagg 100080
aaaaaatata cacatagggt ttgatactgt gtgtgatttc aggcatttgc tggacatctt 100140
ggaatgtttc ccctaaggat aagggaggac tgctgtaacc ttgattttac atatgttaaa 100200
ctgaataaat ctcaaaaaca ctgtgttgga ggaacacata cagtatgata ctccttatat 100260
taatttttaa aatagagaaa ataattatga ttgatatctc catatgtagg aaaattaata 100320
aataaagtga attaacctcg acccaagcgt caggtaggga atggcactgg cagctcctct 100380
ttagccttac ccgtaatgca ttatttctta ataaaaactc tatgccaaag aatatatata 100440
tatatattct tatgtatata tagaatatat acatattctt tatatatgta tatataaaaa 100500
catacacata ttctttatat atgtatatat ataaaaacat atacatattc tttatatatg 100560
tatatatata aaaatatata tatattcttt atatatatat atgttgtgtt tatacattgg 100620
tttgttattt catccaggtt cctacattct ttcttggtgg taacagctca gtgacttcat 100680
ttgattcagg tgaatgcaga ttggacggaa gtttgcgtgt tctattcaga atccttcaca 100740
tattcaggac tttgacagat tcataggtca gtgccttctg gagcttgtcc aactagagaa 100800
gttgctgtcc atgcaaaatg gagctgctca ttaggctggt tcattcatgg tccagaccac 100860
tggctggaat ttgacctctt cacaggcaag accactccac tttctctctt gggctgtttt 100920
tcctctcccc agtctctttt ccaattacat tctcagtccc taaatcttga tttgcgtaag 100980
taaatatatt gtttccttgg ttattaatgc aattctccta ctctcctgag aagctcagca 101040
catacgggtg gtctaataag cacacccttc tcaaggagag agctgggtcc agcatgtggg 101100
gaaatggtag acaggaaaca aagtcctagg tgtctgtggc tcctccacct gacccctttcc 101160
ctgctgttca gctttaaaaa ggatgattgt gccaggatga aggaaacagg aagcttttgc 101220
aaaatcaata ggagggcttt gctcattggt gtaataatgg tgtaacatag ggaggacctg 101280
tggtaccaaa tagtagtcat attatctcag gaaccagagg attgcttttt ttttttttta 101340
tgaggcctga ttctttcagc ataaaaggca tgaaatttaa agacatgaaa attactgaat 101400
ttcatattat tttcattact aaatcctcct tttgactgtt aatgatgctt ttttttttg 101460
agacggagtc tcactcttgt cgtccaggct ggattgtact ggtgcgatct cggctcattg 101520
caaactctgt ctccccggtt caagagattt tcctgcctca gcctcctgag tagctgggat 101580
tacaggcgta tgccaccatg cctggctaat ttttgtattt ttagtagaga tggggttttg 101640
ccatgttggc caggctggtc tcaaactcct gacctcaggt gatccatcca ccttggcctc 101700
ccaaagtgct gggattacaa gcgggggcca ccatgcccag ccctattaat gattcctata 101760
gtgtaaatgc atcataactt gggtcatcca tttgtttaat gtagtaactt tcatttataa 101820
aacatgttga ccatagctgt taccttttggt tttcctgggt gggtaacata ttaatttttg 101880
cagatatgat ttatgttctc tagaaattaa accctgccaa ttttcctgtt attctttaca 101940
ttcatcttgc actattggca gagtttttgt tgctacttta aatctttcag tgtttttcaa 102000
gaactaactt gacagcattt gtcacacttt tttcttgtct cagtcactaa gtagcgtttg 102060
ttcctgtcag tgaatttcta aacttttaac aaatcagaaa aataacactt tcttttcttt 102120

```
ttttttttatt ttttttgagt gaattcttgc tctgtccccc aggctggagt gcagtggcac 102180
gatcttgggt tactgcaacc tctgcctccc aggttcaagc cattctcctg cctgagcctc 102240
ctgagtagga gtagctggga ttacaggtgc cagccaccat gcccaggtaa tttttgtatt 102300
tttagtagag atggggtttc aacatgttgg ccaggctggt cttgaactcc taacttcagg 102360
tgatccaccc tccttggcct ctcaaagtgc tggcattaca ggcgtgagca ccgggcccgg 102420
ccagaaaaat aacatttttct aaaactttat tcctatgttt gaactctcaa atgtttctga 102480
ataccaaccc atctgtttta agtgactact acaatggttt ttggcttatg agtgtggttt 102540
tcattgtctg ttttatggca gtgtaatacc aaacctacaa tacaagaaag gtctcaaagt 102600
agaagatgac tcattttaat ttgatttact aaaaaaggcg gattaactca tttgtgttta 102660
taggtgttgc tatatattaa tggaatcttt tttaaaaaga cagctggggc cgggtgtggt 102720
ggctcacacc tgtaatctca acacttctgg aggctgaggc gggcagatca cttgaggtca 102780
gcagttcgag accagcctgg caaacatggt gaaaccctcc ctctactaaa aatacaaaat 102840
tagccgggtg cagtggtggg cgcctataat cccagcactg gggaggctga ggcaggagaa 102900
tcgcttgagc ctgggaggca ggggttgcag tgagctgcga tcacactccc ttctggacaa 102960
caaagtaaga ctctgtctca aaataaataa ataaacaact ggagactgtg tctctaaata 103020
aataaataat aaatgacagc tggaaattcc ttctttgaac attaaattat tagttggaaa 103080
tatttctata atctatatta ctgttgtggt tgctacttgg aatttttaac tttttacata 103140
aagcaaaatg taattaaacc atctctctag tatccagcaa gcacaaacgc aggagagctt 103200
gctaagaatc aaatatcccc tctccttgcc agggctaggt cctgaggaga cacagttggc 103260
ttgctgacaa gtctagctcc atatcatatt ctcacttaaa acttagtcta aaaaaagtga 103320
aaaacacatt tacctatatc aagctagtgt gtctacatat gaaattgtgg acatcgttac 103380
aaatcacaat ttgtagtcca aattgccagc ctttccctct atgaaatcat tccttgccaa 103440
tacaaatagg aagacagaaa gtcatcccta cctcctgtta gcatttgtga acatttgcaa 103500
atacatttgt cgttgtctcc atcctttgtg ctaaaatcat ttcctggttg gctgatgctg 103560
cttattttgc cggctgtccc tgtaagtcct ttraggtgaa tcctgtaagc gtgcaaagaa 103620
aaaaaacaca ttggctaggg tcattgattt accgtagtgg caaatttttt gtgatgaaga 103680
attccattct acagaagcgt gttctgtact cgttaatgga ctaatgcata ctctggacaa 103740
aatattttgc actggtataa acaggaacca acttatcatc aaatccttca gcaaagaggg 103800
atgttttcat gaaaccttca acacatatca cttgcacaac tatcagaagc gactgtagag 103860
ccctgtaatt tattttcctg ctgctttcag ataaacagaa gagaaagaaa tgcagcacca 103920
ggctcctcct cccaggtctc cagtcatctt ccatagagac ggagtcctga gacaactggg 103980
caacctcaaa cattattttc cgcaggggcc ccgggggga tggagaatgc agcagacaag 104040
gaatggccac tgagtttggg gaagaaatct acagaacggt gctgaaaata aatccttgtg 104100
gctacatttc ctcatgtctg tatagtaggg taatgtaatt aaacttttag acattgagaa 104160
aggaacaaat gtcggagtaa gttagacact atttacaata cagacgatcc ctgacttccc 104220
atggggctat gttctgataa gcccattttc tgttgaaaat gttgtatatt gaaaatgcat 104280
ggaatacacc tgacctttgg agcatcatag cttagctctg gccttcctta aatgtgctcg 104340
gaacactcac attagcccac agtcagacag agccatttgg caacacggtg cacgcagygt 104400
ctgttgttca ccctggggat cacaggactg actgggacct gtggctcgct gccgctgcct 104460
ggcatcatga gggagcatcg tgccacatat cactagccag ggaaagatcc aaatttaaat 104520
cccaaagtgt agtttctgct gaatgcgtat caccttcaca ccatcgtaaa gtcgaaaaat 104580
cttaagttga accattgtaa gctgaattgc aaaaatacgg cttacatcgg tcatctgtgt 104640
accagcaagg agcatataag ggaagggaga agacaatatt tttgaggttg ttttttcttt 104700
ttttttttt tttttatttt ccataactat gctcaagagt ttctgctgca aagaagcttc 104760
ttggcagatg gttcaggaca gatcagagca ggcattcacg taatggggta tgccatgttg 104820
gcacgttggg tcctcacgtc ctgatggaga aacaggcaca cgaagaccca ggcgaggagc 104880
ctacaaagca aatcctgcaa tggtggcagg agaagtgtac ttgaagcacc aagatgatgc 104940
ccttctttgt aaaacctgct aatgtttgca agctgccaca ttggaataat ataatttcta 105000
acagtttgta ttggaagaat acaaagaaga gagaaaatgt tcttttagtt ttacctgctg 105060
gtcgggccag gccaggtgct tacacctgca tgcacactgg atgcttataa ccacgtgcag 105120
tggtggccgc catctttgtt ttggcactga aagtcactga ggttcagaga tataaacttg 105180
tccgaggtca gactcttaag tcatggaggt aggatttgca ccagatgcag caaatgcctc 105240
tgccatgttt caacactggt gcacacctaa acagagatgt ttgtttgttg aagaagttgt 105300
gaaaagatga gggtagggcc atgtgatgtg gagttccgta agtgttgctc ctaagtgact 105360
tcagtattaa ggcagcccta gaaacttcat cctaaggcat gaactggaca tgtgagtctc 105420
agtattttcc cacacgtttc aaaagtgaga ctggccgtag ctcagtctct aaatgcctgc 105480
tgcaaaatgc taatgtcata aatactcatc tctgttggga ttttgaaaca ctgtactttc 105540
tttccattgt cttcccatta atcatagaca ggattgagat gaaccacttc ccttgcttat 105600
cttttaactc tctcttgtct cctttgaaca tgtttagttc tcatggaact tgttaaatta 105660
tccccagagg caagaaaaat aagggagaat actatttttt atgagtctct gttagaaagg 105720
ttttgtgtaa ttttaggtcc ttttgtggcc cactggttta aagtgctttc tttaaaattt 105780
ggttattaag aatggccatg ttcttgaagt tgctttacat tggtatgggt tgatttttt 105840
ttttcaatct ctgcagcttt gccagggatg attttatata acagtggagt aaagaggtaa 105900
```

catcaacatt aacaattaaa cctcagtgtt atataaaact gccagaatgt gtgtgaaaag 105960
tgatgaattt ttaagattta atgtacgcat agcttttagt ttcactagaa agaatgaaat 106020
tctattgatg catttatgca tttcttataa catgtatttt ccagttttcc aacacttggg 106080
gaacatttct tctgggaaaa aaaaatccct tacatgctgc atacaactgg cgtctcaaag 106140
catttgcagt tatgagaagt ttcagtccct tcacagttct cttatcatgc tagcatcatg 106200
ttttattagg ataataattt tcgatgtaat atctatttta tcttgccaag caaattaagc 106260
tttaaaccaa tgtgtgtgtt tttctaaatg gcctatccaa aaattgattg catttctaaa 106320
ggaaatatct gttaagaacc atctcagttt aaaatatttt tataatgtca gcrtacaagg 106380
gtaatgaccc attttgtaaa aatcttytta tacaaacagc ctaatcctta atttttgtgc 106440
ttctttttt ttttttttt taattcttct gttgtagatt cctaactgtt gccagttgaa 106500
aaaatattta acttggaggt aaaacactga ccaaccactt gtgtctcaaa attcattgaa 106560
gttttgatct ctttggagtc aagttggaac tgctgtgagg cccaaacacc tatcttctca 106620
ttcatctcgg ttgttgcctc tccaggagag cctgatcttt ccataatgag aatagtgaat 106680
atgcttcact gatgtttaaa gagtcacatc catgtatatc tgtttctcaa acatgcttct 106740
gaattttcat ccactgtttg tacagcagga catactgggc attgtagagt tttcagttgg 106800
ttgttcaggc aacttgacat ttagccgctt ctccgtgctg cccaccacaa tcctcccctg 106860
ggcagcctgc tcaaggactt taacattgtg tctcctttca gactgttcag gtcgtggagc 106920
ggagtgtctg acttgggcat taatgagatg aagacgagac tgtaggtcag atgatgactg 106980
tttttgtgat gttcgtgttg accttcattt gctaatttct gacctcaaag tgggtatttc 107040
ataatgtgtg ctccatgatc acgaggcgcc accagtctgt gctctttaga ctcctttagg 107100
ctggcgttgg tgccagtggg cacacagtct cacttctctg cccctcccgt tgcacacaca 107160
tttcggagtg cctctatgtg ccttgtgtac cagcattact gtgcatgtgg cttcaccgta 107220
cttatcttgc acactaggtt gtcaagtccc attgctgttc tctctctcta ctctcatggc 107280
attttagagg cagaaagtaa attcccagtc aaggttgccc atgctattac ttatgattat 107340
tgctgccaaa tgggtgagga caaggtaaac acccagggaa tgctgtgaat ctgatgtatt 107400
tcctgtagag gagagcagag ttgactaacc atcccaccta actctgccat ctctaaactt 107460
gacaactaat cttgactttg agattgaaga caattgaatg tgtttaaact tcataaagac 107520
agactaactt ttgaaacctt ttggaataaa acagcacagt cacaagtatc catcatttat 107580
gctattcatg tgacatatta tcatgggaac acttactatt cactgatttt acaaatacct 107640
atgaaagcca attatctacc aggcagagtt cttctaggct ttgaagatac acagtaaaca 107700
caatggacaa aatactgttt gtatgaagtt tcttttatat tgttataacc aaagttagaa 107760
ttttaaaccc agagaaactt aaagaagtaa tagtttagat cttggttaaa tcattgtgtt 107820
tctcattttc tggaatagtc acccagcaaa ccttttaatt ttttttttct ttttcttttt 107880
cttttctttt ttttttgaga cggagtcttg ctctgtcgcc caggttgcag tgccgtggca 107940
cgatctcggc tcactgcaag tccgcctccc gggttcacgc cattcttctg cctcagcctc 108000
ccgagtagct gggactacag gcgcccacca ccacacctgg ctaatttttt tgtacttcta 108060
gtagagacag ggtttcactt tgtttgccag gatggtctcg atctcctgac ctcgtgatct 108120
gcccacctcg gcctcccaaa gtgctgggat tacaggcgtg agccaccgtg cccggccaca 108180
aatcttttaa ttttttcttt caattaccat gaactcactt acctataatt gagttcttca 108240
cttgagagat agaaatgttc atacaatgag taagcctcat tcccttccca gtctttaagg 108300
tgtattttaa gcacrtagcg ttgctgrtta gtcagttgcg aaacaaactc atttcccagc 108360
caatattctc ctgaagggtt accaaatccc tgtaatgcaa gttgttaaat tcaattattt 108420
catgtaattt tttctttgta tatttgaagt ggatagtccg tcaacttaac ayagaataac 108480
tatcaaatag cagaaattcc ttctggtgct gtgacaattt agggtccttc ccaaaggaaa 108540
atggatttta aataggtcag ttattagata ctaagctgct gctggaagaa aacttgtatt 108600
aggataatga gaactacttg gggagccacc agcagaagcc ttggcataaa cagctcagtt 108660
catgggaatg tgaagcacca ttaaacagtc ggcttaccaa aaaaatgctg agtccacctt 108720
taaaaataag ctaagtagtg gcagccttgt ttatttgaga gtcttactct gttgcccagg 108780
ctggagcgca gtggtgtgat cttggctgac tgtaccctct gcctcccagg ttcaagcgat 108840
tctcctgcct cagcctcctg agtagctggg attacaggtg tgcaccacca cacctggcta 108900
atatttgtat ttttagtaga tacagagttt tgttatgttg gccaactggt ctcaaactcc 108960
tgacctcagg taatccatct gcctcggcct cccaaagtgc tgggattaca ggcgtgagcc 109020
agcacgtctg gctgcagctt ttgttttgat acagtttacc ttatattggc cattctttaa 109080
aggggagact gaagcaccaa ttttaaaaac catgtcaaaa gtcattggtt agtttgggat 109140
tggtggttaa ggttccgcag atcttgaaag ctatttttca caagggaaat tctttyctga 109200
tgccttaaag aatgtcctta ccactttata ttctttccaa gtcctctgaa aatcaacgct 109260
gccatcctca cgtcgctgaa taattgtcca cccgcctcct ccagcttcca tgtcacagta 109320
ggcctgcaaa caggaatgca gggtataagt gacagagccc ccccactccc cccttacgta 109380
gcagaagcag gaggaatgta gaccctgagt gcaggactca gccgagaggg ttctctggga 109440
tataaggcac ggagtagacc atcggggttg tctaagaaac agatggtttc aaataaattg 109500
aaagtgatgg attaaaatga tggtaaaaca taaacagtaa tataatataa agtgttgatg 109560
agaatatgac tcacttgtca tcatctgttc caggctaaag accccaccct gatggctggt 109620
ccaggagatt gctgtttttt tagagattca ttatgaatgg cacattttgg cagattggcc 109680

ccgaacccca catctccatc ctgtagaaaa ccattgactt gtgttcacag ccttgaaacc 109740
tttcactaaa tgccagcccc tgcccctcca cagagagcta tgtgaagggg atactctttg 109800
atcatagggt ttggaggact cctgttacat tctcgtactg tggggctgtt tggcttcctt 109860
tatgtgcaat actaatcaga ttttttggtt cacatctgag cacaagggcc ttgaggctcc 109920
agtgtcctgg tgcaaggtgt gtgacctggg cttggcttag caccttggcc tcaagggcat 109980
gacttcacgt ttctctgtac atagctcact gcccacagtt tttctctaag caactctttt 110040
tatttctgct cttaaactgg ctctgtgggt ttaaactttg tccagaacac agaattcttt 110100
cttaagctaa gcgcatgcat gctcaatttc aactcagctg gagttttatt atgaaattaa 110160
aaacccctga aaataagatt acataaatag ttttaaataa taaagattac agtagaacga 110220
aacatgtttt ccagaaagta agataaattt tctgccacat acaaggtagg aaatattgaa 110280
gttaaggttc aaaaccaatt gtaaatattc ttcttgtgtt gtgtcccatg gtctttttga 110340
gaataatgga ggcgatttgg cagggtaagc ttcaacgcac actgttctct gtttacgagt 110400
tagaatccta aagaggagag cgaaaagaga actaagagag tgttattcct gttctttcat 110460
ttcctacatg aggaaagagg ggtgcagagg aagtgcagtg gctgcctgat gcctcattgc 110520
cagtgcaacc agagaactgt ccctggaaca gcctggtctg caggggactt ctccccagca 110580
tggttctggt ctctctccag gaaggtcacc cggctcgtat tgccttcccc agtggcactc 110640
atgtttggaa agataggtgc cattaatgaa agttacattt attaagaaga aaattatttc 110700
ctcattttaa attaatctga tgtagaaaat cagtctgcac aagctaaccc cttgttaccg 110760
ctctgtattg ctatgatttt tattatcatt gctacacacc actctcatcc aagtactctg 110820
ccagatactt tgtgtatatt atctataatc tcacagcaac tctgctgtag tatcataatt 110880
ctgcattttt gaaaggagaa aagatttagc aaagttcgat tttgatcagt tactcacctt 110940
ctaagtgata tataggattt gaataaggtc tctttgattc ctgttatgtt ttttttccac 111000
tgacatcatg ctgccaaaaa tagagaaacc tgacccttg gtaaataggc caaaagtccc 111060
tcaacacagt tccaagttta tatcagttca tgaataatac tgcctcttat ttgcctgcag 111120
tcaacaaaat ggtcagtgct gctcacttct atcaatattt ctttttaaaa atctatttca 111180
taaaccagct gaataagcta ctttggtttg aggattcatg tacatatatt gaagttaatt 111240
ctgtatccta aatggtatgc tcttgggttg aaagcattga cgtggctctt ggtgagccta 111300
tccttgttcc aagaatgttt gattcttcag cgtcaaaatc actgctatga agttaccagt 111360
acttaaatac atgttctgtc ttgttcagag aacaagttta tcttgttatg gaagtcagag 111420
gcaaaactct taaatgtctg agagtcactg ccagcacata aatgatttga gccatatgag 111480
tattcgcttt gattccatta gtgatgatga taaggttatt agaacatttt cttagtactt 111540
catccaggtt tttagaaaaa agaacagagg atttgtaaaa actggagtat tatggttaat 111600
tggactataa aacttgcaga gaaagatagt gttcaaatag agttatctac ccagccagaa 111660
gatactgagt aaaagtgctg aaattgatta tatcaggatc agcaaagcag aagtcctcag 111720
atacttccca agaccttacc actccaatta caacaaacct aagggcagtt aatatcttta 111780
atctgtccac tggtgcacgg tgcaggaacc tgatatcttt ctgtaaagct tgatgttttt 111840
cagcaaataa tacttgactt gcttcaactc tgaggcaatg attaagtgac gggttaaata 111900
gcaaaccata gagacaaacg ttaggagtca ggtgtcctgt gaaatttagg gaaggaaatg 111960
accatacatg cttttgataa atgccatctt gcagtctctc tctcgtagaa gaaccaaatg 112020
aatttttcaa aactaaagct gcagtatttt ggcctttcag gaaaagatct gctcaaagac 112080
caattgaaca ttctttttctt gaattagata aatgagtgca gaatcgggtc tcctgccagg 112140
cagagaagtt gtctggtagt ctttgaaggc agcgaaaatg gtgaccacta ccatttactg 112200
tcccaagttc taccccaggg gctgtccctg tgttgtcaca acccccaagg ctgagggtat 112260
cattgttctg ttacagatgg ggaaactgag ctctcagaaa ggttaaatga ctcgcccaga 112320
gccacaaatg gcagagctag aatttacatc caagtctgtc tatctctccc tggatcagat 112380
tgtgacttac agagtcttaa gtctgcaagg gaatttagaa gtaacaattg acaccactta 112440
tttttcaaga tgagaaaatt gaggtttgtg aaagcttatt gttccctgct gtgtaaatga 112500
gtttctttta ctgcaaatgt ttaaagggaa tataaatcct aatgtttcca accatgacct 112560
gaggctcata taatcccaga gtactatata ttttcatctt tctgcagaat atttcagtta 112620
ttctgggggc catggggtgg gacagtgcac tggggtggga agcttgcact tagactgaga 112680
aacatagatg aaacaatgtg atggggctgc atggttagcg gctgtccctc tgcatcggtg 112740
tccatggcat ccccatttca catgcatcct ctgtcccccc tcttgacact cctgtcccca 112800
ggccaagaac acgccaacat ctggtgacag acaatggcat gcacagaatt gtcatgaaaa 112860
ccagatagca aaagagattg aaaggcttag cctagagtct gttcgttgcc ttttcatctg 112920
cagcagaccg tgttgtttgt gggtttgttc ccttccttcc ctgttgtatg gcttctaggg 112980
cgttagggac attaactaat tttcagggtt gatttaacac agcattaatg aattagaaag 113040
gtttctttgg aaagcattat gttgaatagc acaatattta tcttttccgt tcataatcaa 113100
gatacatgtt actgtttcaa gttccaggtc tttaaaaccc taatgcttgt atttttaaag 113160
tgtttttctt tgatactgtt ttaaatactt aggataatac tcaatttaaa gagataatac 113220
ttccaagagc ctcgaaaatt tccactttgg tacagtatcc actgtattct ctgtagttat 113280
ttgtgttggt tcaatatgta gctgctttta catttatatg caaacatatt tatagacatt 113340
taatatatac agttatccag taattacata acacttcacc acactgattc tcctgtaata 113400
tcattcttcc ctatcaaatg tttaagagaa gccacattga aatattctcc ggaagggttt 113460

```
ttttttttcc ttatctaaag ttcagtgtct cccaaagcac cttcaggagt caggctctct 113520
gagtgaggct gcagaactag gaatgactga ggtaagctgt gttgtggctt tgcctgctgt 113580
gagtactgac tgtccactca cgttacatgc agtaattgga catatgcctt gaagtgaact 113640
ccgctgctgg agaagaaaat gaacactgtc tttatggtgt gtttcgagtc ttccagactg 113700
cgttagatga ggttagaatc gccttcccca gcggttctca tggtgtggac ctcagaatcc 113760
tgcgtgtccc tcagaccatg tcaccaagtc cacaggttga aactgttttc acgatgctaa 113820
gacaccaccg gtctgtggca ctgtgttagc gtttgctcag atagagcaaa aaccatggtg 113880
ggtaaaactg ccaccatccc cgtgtgaggc agggcagcgg tagctaactg tacaagtcac 113940
tatgcagtca cacacaaaaa agaaaggaac aataatatca ctgaaaaatg actttgacac 114000
agcagtagaa attattaatt tcatgacacc tctatggctg atgcactatg gttgttcaac 114060
cttcagtttt tggaagatac tttctttaag gaaatgagtc tgacacctca ggaaaaacaa 114120
ctgacagtat gcattattca agcaaaatgc aaggagaatg tatttgttgc cgaagataag 114180
atagacattt ctacatcatt catatgcagc ttttgtattt tctgcacagc agcggcactt 114240
agagcccttg gcaagcactc taggcgaggt agctgcccag taactatggc tttgtactgt 114300
gtgaggctgg ggaagatctt tggggagaga cttctgctct ctggttcctc actctctaaa 114360
gtggcctgcc tagagccagg gagttagtaa ggggagacga atacctcacc ttgatctctt 114420
ctgtagaatt agggaatgtt aacgtgtaga tgccattcgt ggtgtgtcct gatttgaata 114480
cttcagcaca gtctctgaag ctgatttgtt cttctttagc aacagtgggg tccttagctg 114540
cttttaaaaa atagtaaggc atttaaacgg agttcatgaa aagacaaaga cttgttattt 114600
caasagccaa tcatttggtg agttttatta ctttggaatt cttaagtaag caaaaggctg 114660
taccactttt ttaacctttc tagaaagttt cctttcagcc tgttttcttc ttaattctca 114720
aaagattaat acttactttt tggtcattaa ttccatgtaa ttaaaatact tcaaataatc 114780
caaacttcct ttgctgatga tcaattacat gtaatgaaag tacttcacaa tcacataaat 114840
taaattatta cttttgaaga tctttcatct tgagtagaat agggtaaact tagtatggaa 114900
gaatttaaaa agaatgtccc taaacactgt tatctgtatc atgaccccat tgcctgcccc 114960
ttcaggccat tatcccactg aggacatagt ggggtgcagt gacacatctc agcttaccgc 115020
atcctcctcc tcccaggttt aagtgattct cctacctcag cctcccaagc agctgggatt 115080
gcaggtgccc accagcaagc ccagctaatt tttgtatttt tagtagagac aggatttcac 115140
catgttggcc acgctggtct ccaactcctg tcctcaagtg atctgcctgc ctcagcctcc 115200
caaagtgctg gaattagagg cgtgatccac catatctggc ccttccctcc aatatataag 115260
agacgctgca aagtgaaaca ataataagga aggcaaaatg tgcttaagaa cctggcaaga 115320
taagggaact agcatctctt aagtgccagt gtattatctc atttaatctt aatggccatc 115380
ccatgggtct gatattattt ttcccatgta aaacctaaat aaatgaatat cggctgtggt 115440
ttagtaattt gcccagtctc atccttctaa ttaatgatgg aactaactaa aagtaggctc 115500
tttactgcca tgaatcaaaa gtatgcttgg ggtgtttgct tcataataat tagtataaca 115560
tatatttccc cttctcttct tccttcattt taattggtag atatttcatg tgaaatatat 115620
gagaaatagc gccttttctg aaaggtgaga atttttagt cttttgagtg ttttactgac 115680
taaaggttat taacgctgaa gaaagcatga tatgtraact tacagtttga tgtggacatc 115740
atagtcagta agttattaac tgtctccatg agatcatgtt gctgcttctg aagaactgaa 115800
ttattcaccg tggcagtcac tatttttttt tctagttctt caatgatgga attttgcttg 115860
gatactaaca cctgtagctg atctttctct tcttttattg actgtagttg gatgatgtgc 115920
ttgtcttcca tagctagcac cttcttttct aggaaacttg taaggaaaag aattgttagt 115980
tagtgaaggc tattctaatg aaatatttta tatttattga atttctactt ctccaaggta 116040
ctctgttaag atattgtagt ggttataaag taatatgatc ttaccagagc cctaaggaat 116100
ctctgaaact tgctgagaag attagatata taaatgtgtg tatatatgta aacgtataag 116160
catatatgta tgtacatgca gacttatgca tacacacaag aaaaggtacc ccatctggtc 116220
caggataggt gggatatggg tgtttttttgt attagatgct acagcgctca gaagaaaggt 116280
gctgctcttt caagcttagt gctcatgaag tgctttttttg agaagggaga gtttcaactg 116340
ggctggaccc ttgggtagga tattagcttt ctcctaaact atttatattt taatattaat 116400
cctaatgata ataatagcac ttaatgctat gtgagaaata ctccttcatg gggaggtgaa 116460
tacttctccc agactcaagt cctggcttac cagccctgcg acttggaaca gtttacttag 116520
tcaccctatg cgttaatgtc ctcacctgtt aattaggata ctatcaccta cgtcatgggg 116580
ttggtgtgag gaacaaatgg gttttaaaat gtaaatgctg gccgggcgca gtggctcacg 116640
cttgtaatcc cagcactttg ggaggccgag gcgggcggat cacgaggtca ggagattgag 116700
agtatcctgg ctaacacggt gaaaccccgt ctccactaaa aatacaaaaa attctccagg 116760
cgtggtggcg ggcgcctgta gtcccagcta ctctggaggc tgaggcagga gaatggcgtg 116820
atctcgggag gcggagcttg cagtgagctg agatcacgcc actgcactcc agcctgggcg 116880
acagagcgag actccgtctc aagaaaaaag aaaaaaaaat gtaaacgctt agactagcgc 116940
ctgtcataca ttaacactca atgaatgttt gttaacgtta atatagacat tattattccc 117000
atttccaatg aggaaattga aacttaggga cattgagggc caggctcagt ggctcacacc 117060
tataatccca gcactttgga aggctgaggc aggtgtatca ctagagtcca ggagcttgag 117120
agcaggctgg ccaacacggt gaaaccctct ctctactaaa aatacaaaaa ttagccaagc 117180
gtgggggtac atgaatgtaa tcccagctac tcaggaggct gaggcaggag aactgcttga 117240
```

acccgggagg tggaggctga agtgagctga gattatgcca ttgcactcca gcctgggcaa 117300
aagagcgaga catcgtctca aaaaaaaaaa aaagaaaaga aaagaaatat aggaagaatg 117360
aatcacatac ctaaagtcac acacagcagg tggcaggggc agaatacaat cccagcactt 117420
tctgactctg aaatctgctt ctctcctttt aatgtggccc cattccttct ctaaaaaatc 117480
taaccagcct atcgcatgta cttaatacat aacagttaat atgtgagcca agcccttgaa 117540
aagctttttt ttctcttttt ttgagatgga gtctcgctct gtcacccagg ctggagtgca 117600
agggtgccat cttggctcac tgcaaccttc acctcccagg ttcaagctat tctcctgcct 117660
cagtctcccg agtagctggg actacaggcg catgtcacca tgtcaggcta acttttgta 117720
tttttagtag agatgggctt ttaccgtgtt agccagaatg gtctcgatct cctgacctcg 117780
tgatccgccc gcccctgcct cccaaagtgc tgggattaca ggcatgagcc accacgcctg 117840
gcagaaaagc tttttaaaaa ttatttagag agctggtaaa attatgccat gtaagtccta 117900
agacacttta ttaatggtta tatagtttgc cttcctaatt tcaacttata aacatacgtt 117960
gctataaata tgttcaatga agagcatacc acttttaaac taaaaatagt tcctgtccat 118020
taagccagag gaaacaaatc caagagagta gagactatgt atttgagaat gttaactgtt 118080
tcccaggaac aaactcaaag acatgcacag tcaaggtatt tggcagggtt ttttgttttt 118140
tgtttttgt tttgagatgg agtctcggag tctcgcgctg tggcctgggc tgttgtgcgg 118200
tggcgcgatc tcagctcgct gcaacctccg cctcccggat tcaagcagtt ctcctgcctc 118260
agcctcctga gtagttagga ttacaggtgt gccaccacgc ccagctaatt ttttgtattt 118320
taagtagaga cgtggtttca ttatgttgtc caggctggtc tcgaactcat gacttcctga 118380
ttcgtccacc tcggccttcc aaagtgctgg gattacaggc atgagcaccg tgctggctgg 118440
catttttttt ttttttaata agatacaaga ggaaaattgg atagcctgac actacattat 118500
tcagcaccta aagaggcttt ctgtgataat tgcaggaaaa gcagcaacta aagatgtttc 118560
aatatcttca ttttgtttgt acaaggccag taaataaagc tttcaaaata tagacacttt 118620
taaaaataga aaaacagtga ccagatgtca gattcctctc tctgacattt tccttccaat 118680
ataaagttta gtacacatga atttgcacat tgcagagttt tgttttaaag gaaggggacc 118740
tcatattccc ttttttgagt cccgtataag tcagctatct tatttaataa tgaaatatgt 118800
caatgatggc atctttatgt ttcagaatta ttttctgtct actaacaagt taccacagct 118860
tctgttaatg tcacattaga agctggtgaa atattctata catttcacta gcttttctgc 118920
gaaggcatat gaagagcaga gaaacattat tttcccacct gcttgataaa gaaaccttga 118980
accggccatt taacactgct gtgagttatc tgaagcctcc tgagtcactt tgcacttact 119040
ttcctaggaa ccgaaagcat gtgaaattga catacacgtt tcactgagtg atagttgggt 119100
tcagatcacg tcttaccttc cgtttaacag agatgtattg aacacctacc atgtacgagg 119160
tgtttttag ggttttggag aaaaatcaag aaatgaaagc atcatgaacc atagtcttaa 119220
gcctgcggaa atttagatgt tttgatggtc ttcacatcat caagctaaaa agacaaggct 119280
atgaatgtct cccttgagga aaaactaccc ttgtggccat gtaaggtctg taaatagaag 119340
ttatcacagg gaatacatat gaagatcatg gtttcactga agagaaaatg gagaccctga 119400
gaagtcacct ttggtgtcac gagcaccttc aggtgaaagg aaggagcctt aggctgggaa 119460
tcccacctct gcacatggct tcctgtgtca catgggcagc caccctgctg tggacctcag 119520
gtgcatgtct gtctaggtga atatctatct aaataaagct ctatgtaaaa tgaaggcatt 119580
cgatttcatg gcctctcggg ccccttttag ttcgaatgat ctggtaaatc caccttttt 119640
tgacagtaac attttctgac tctttaaccc tgcaaacaat attaaccagc caaggaactg 119700
gctacccatt acatgtctcg cccataagca ataacaatca gtattaataa taattattag 119760
atattcaatt gtagctctta aatgtattcc agccccctga tcgttgtaaa ttagtatata 119820
attttggaga gatgggggtc tgtctttgtt gcccaggttg gtatcaaact cctaggctca 119880
agcgatccac ctgcctcagc catccaaata ggagagatta caggtgtgtg ccaccacatc 119940
tggctaattt ttgtattttt tgtaaaaatg agctcattat gttgccctgg ctagtctcaa 120000
actcctggcc tcaagaaatc ctatcactct ggcctcccaa agtgctggga ttataggcat 120060
gagccactgt gcctggcttg aatctattct ataaagaaag caattgcact tttggggaat 120120
tataaaagat tatttaaaat gtggtttgtc caatgtgaaa caccatttgc atatttttgt 120180
aatgatatac ttgcaaataa aatcataggc cagtcagaat ttaaggtaga aaacacagca 120240
tgcagaactc atacacctgt aaaatcatca acactatttt ctttttttat tatttatagc 120300
tgttgatgaa aaaaaccttt ttatttcctt tcatatctgt gacaaaaaaa tacgatttct 120360
acatctgatg agaaaaagct tattcttcct acaggcatag ttgaaagcca atatgattgg 120420
aaaactattt gcaaagatga tatttggggg acataattga cccaaattgg tagttttagc 120480
attgtagcat gctaaatttg aaacccaagt ggggaaacag tattcagtat tagggtatgt 120540
tctacaaact ggacatatcc taggtttgtc acggacatca ttgtataaca ggcaagagaa 120600
aagtaatctc cagctcccat gtgttccggg aatcactgca gcattttgaa gagaacatta 120660
ctaagtaaga ctattaagaa aacgacgcca ggacggtggc tcatgcctgt aatcccagca 120720
ctttaggatg ctgcggtggg cagaaggctt aaattcagga gtttgagacc agcctgggca 120780
atatggcaaa accttgcctc tactaaaaaa aaaaaaaaaa aaaaaaaaaa aaatcagctg 120840
ggtgtgtgac acatgcctgt agtcccagct actcaggagg ctgacatggg agaatcacct 120900
gagcccagga ggttgaggat gcagtaagct gagatggcac cactgcactc cagccagggc 120960
aaccagagtg agaccctgtc tcaaaaaaaa aacacagaaa agaaaatgaa attagcagga 121020

ttgttatatc tcaatgattg gtctcaaatg ttcatttact gtttgtagag gagaaatctg 121080
aaacatgaaa gaaaaatatt tgaattttaa aaatctattt gcttttcaaa accctaaatc 121140
aataatgact taaacttggt atcctaagga cagaaagaat tatttcagct tagttcttga 121200
ttaacagtaa agaacaatta ttgaacaaga agtttatcat ttttggttaa gaataaagaa 121260
ttatttaaat tgtcaaatag gatatattgt tatagccatg ttccatgttg tatatacatg 121320
tcttcattaa aaacaaggaa ataggcacac caggtatgtg cataaaatta tcctcttttg 121380
tcccaagtgg aacagacata tgaaaacagt ccccacctat cccctacaat tttttttcta 121440
ttgttgatct tgagattttt ctatatttta tttaaatatt aatataatca tgtttaatat 121500
ttttggtttt actttatcgt gtgtttgaag aggaaacatt ggatcataaa atgtgcattg 121560
gcttacagta taagtgtagc tttcatacta tagaccattc tgcgttgagt gaagctaagt 121620
ccccaagggc aaaggatctt ggtcaagtta atactgaaat aaaatgcctg ggccagtggt 121680
tctttcactc cacagcacta gctgtatttt tataatagat tagcatgtag aatactgagg 121740
cagggtttgg aggattactc taagaggatc ttttgggcca gtggttcttt cactccacag 121800
cactagctgt atttttataa tagattagca tgcagaatac tgaggcaggg tttggaggat 121860
tactctaaga ggatctttaa ggggccaggg aatgaaaggt aaaatccagg actgtgttag 121920
gagagctgtg cctgtgcagg aattttctcc aagccctctc ccttctcctc cctcatgagg 121980
tttctgaccc ttacactaga catgaagaaa ctcaccattc tgataattca tcatttgaga 122040
ccgactttca tatctggaaa gtgtgcagtc ctgaattata aawgttttag tactgttatt 122100
acctgttctt atcttgcaat ttgtttattt cactggtctg gtccaaaatc tgtttttcca 122160
atttgtttgt cgagagggag tgttccaaga gctgaagttc aagtctcgtg gtctgattta 122220
atacctaaat gtaacaaaat gaagttccta ttaattattt tttaattagt ttaactttct 122280
aacttccttt tcattaaagt acccaagcta caggaaaaca taacaaaaac attatttatt 122340
aacccaagta tcttattttg gcatattttt cattttcaga aaaggctcaa tgtcttagat 122400
cacatctgag tgtgttaaac ctttttactc ttttccccac gtctctattt ttttttttt 122460
tgagatggaa tctcgctcca ttgcccgggg tggagtgcag tggcatgatc tcggctcact 122520
gcaacctccg cctcccgggt tcaagcaatt cttctgcctc attctcccca gtagctggga 122580
ttacaggtgc gtaccaccat acccagctaa ttttttatat ttttggtaca gatggggttt 122640
caccatgttg gccaggctgg tctggaactc ctgacctcaa gggattcacc tgcctcggcc 122700
tcccaaagtg ctgggattta caggcattag ccactgcacc cggccgttat gtctctatct 122760
tggaaagtgg ttagtagttc tggacaatgg ggtctgtgcc aaatactaaa tgttatttt 122820
ctagtctgcc atatttttatt tcatacaatg agacaagtag gagtagaaaa tggtcatatt 122880
tcataggtcg aaagtatttt ccctttgccg aaaacaaaat gctattctca tatttatttg 122940
tcactagaca gagagattgg aagtcacatg cttccattat ataaaaatat agataatttt 123000
tagcctggga tttcctcatt tgtcaccact tgtttagact tttatttctt cttgccattt 123060
ctccttcctg ttttaaaact tgtttgaacc aatcgaagcc gtatagcgtg agtgtgaagc 123120
ggascctcag ccttgccgtg cgggcctttg tgagctactg cgtggcatga gcagtgcggc 123180
tctcccgcgg attctctagc gcctggttgc ccttcagcag gaagaatcga ytactcactt 123240
cctccatgtc atgcttattc aggatgtgat atcacaygca aatgtcagtc agcattgttg 123300
ccaaggaacc ggggaccttg aaagaatcat tgtttgctgg tgtctttatg tcatttgcag 123360
gagccttggc tggtccacag cgtgagtttc agggatggtc ttatccttag agctggttta 123420
gttcttatca caaaaagtct tctgtgagaa taaagtcctt ggccaacrta aggttttgtt 123480
tgggttttaa tattaacacc tggaatatag atttggccta cgtcttcttt gagtccaaac 123540
attctatgtt ggttatttct aaaaggaact ggaaaattgt gtcctgttta attcataagg 123600
gttataacat gagtaaaatc ccgtggggag gcagggaagg atggcacata agtcatgatt 123660
ggcccagtag taattgtaac cattttcaca tcacttttct ggagagcatc aaaccgctgg 123720
accagcctga aggcgtccat ctgcagggga ctgtaaatta cccaggccag gtaatgatct 123780
ctcattccct ttaagatatg agacctccag ccacccattg ttgctcaatt tgatcgtctc 123840
tcattctgac cggcttggag aatcttgctt ctaatcagaa attttcagat ttgaatttaa 123900
gtctgtttca caaaatcagt aactgctcag caagtacctt caaacagagt gggtacataa 123960
ttcagtttct ttgcggcctt ccttaagctc agccatttt ctttttttt ttttttgag 124020
acagagtctc actctgttgc ccaggctgca gtgcggtggc accatatctg ttcactacag 124080
actctgcctc ccaggcttaa gcacttcttg taacctcact aagcctccca agtaactggg 124140
tctacaagtg cacacaagca cgcctggtaa ttttttttt ttttttttt tggtagagat 124200
ggggtttcac catgttgctc aagctggtct cggactcctg atctcaagcg atccacccac 124260
ctcggcctct caaagtgcta ggattataga tctgagcaag cgtgctcagc tggctcagcc 124320
attttcatgt gttcaattgg gcttcacatg gaaaaactgc ttactttcca tctgttttct 124380
tattttcctg ttatcctgga taacatgata tctagtttca caataggcgt ttttttttta 124440
aatcatatga cgcaacacaa gtacatcaaa tgctatgaag tctctgaccg ctataggatg 124500
tagcaaggtt tgcattgctg ctctgtccta acacttttc attactatta ttatttttta 124560
ttttttaaa tttttgccaa gctcccatgc ttggatctaa ctattatttt aaaatataag 124620
aaatgttata gtttaaaaat gcttatgaga catttttgg atgagctatt caattaccca 124680
tcagtgttag tatcaaaagg tggggcatgt gacttaatca ttactaattt attttaatag 124740
gttggtgcaa ttttgccatt gaaagtaatg gtggccaggt acggtggctc acgcctgtga 124800

tcccagcagt ttgggaggcc aaggcaggtg gatcacctga ggtcaggagt tcgagaccag 124860
cctggccaac atggtgaaac cctgtctcta ctaaaaatac agaaatttag ccaggtgtgg 124920
tagcctgcac ctgtaatccc agctacgcag gaggctaagg cacgagaatc gcttgaactc 124980
gggaggtgga ggttgcagca agtcgagatc acaccattgc actccaacct gggcaatgca 125040
gtgagactct gtctcaaaaa aaaaaaaaaa aagtaatggc aaaatctgca gttacttttg 125100
gtccaaccta ataataattc gctttagata tatattgata tattgacttt taaatcttta 125160
gtttttatga cttcctagga tttaaatttt tagtacctta tgatccatta tgtaaaatat 125220
ttatgtatgt ttttcctgaa ctgttgtgat attgtggaaa gacctggtaa tcaagtaatt 125280
tgttattcta ttctcttatc tgtaagtctt ttgttaatct atcatttcgc tactgtttc 125340
tctgacctca tccaaccatt tttaggaaga caatgaaaga acagctgtgt ccttctagaa 125400
tgagtcttac gagagtggca gggcttatgg catctcccct ctcatgtcct ctcctggctg 125460
atgtctagca tttcttgatc cttttagctg aagtagcatt taggaataat atggagtggg 125520
gattgtttca cttaaatctg ctctttttt taaaagcatt ccttgtagcc cagagtagga 125580
agccactgac ttcagaagca tgtaaagaag ccaggatgag gagtcagaaa gcgggcttgg 125640
ccgccgagag tcacgaccac ggctttgagc ttggagcgtc tgcatttgta ctgctaatag 125700
cagcttttcc ctttcccacc caggccgttc gctgggtcac atgttgtgca tcatttagca 125760
tgtctctcgg tgaattttct tcttttgaaa ttttcctatt ttgctgttat tttactagtt 125820
tctttctttc tttctttctt ttttttttt ttgagttgaa gtctcactct gttgcccagg 125880
ctggagtgca gtggcacgat ctcaactcac tgcagcctct gcctcctggg ttgaagcaat 125940
tctcctgcct cagcctccca agtagctggg attgcagatg cccgccacca cacctggcta 126000
atttttttgt atttttagt agagacgggt tttcgccatg ttggccaggc tggtctcgaa 126060
ctcctgacct caggtgatcc acccatctcg gcctcccaaa gtgctgggat tacaggcgtg 126120
agctactgcg cctggccact agtttactat ttcagtcttc tttctgttat tattaatcac 126180
tagctcatag aatctcacag tggaaagaga acttagcaat cacttgtctg gcccaaccct 126240
ttatattatt tgaggcccag aaaaggtgag tgcctcattg tgatgcattt atttggttag 126300
tggcagacct ggagccatgg cagcgctcag ggctcttgct cgggcgtgca ccatcttttc 126360
tgtggctaga cgcttctcac tgtcccactt gtctccttct ccataatctc attccacagg 126420
ctgtgttagc tgttgagatt caggtttcat cttaactcaa gagttagatt taaggccaga 126480
gtttctagct ctttgcctca gtgcttttca tttctcaaat gttcaaagac tttaggactt 126540
agaaatggaa aatgattccc ggagtccaga aagcaccagg gagacagagg gggtattcat 126600
cttgcagtgg ttgggatgcg tggcatgaaa atgactcaca tgtcttcagt agatagaaca 126660
catgaaattt aacctcagta ttaaaaacaa aaacagattt actgattttt aattcataag 126720
cagccataca tccttaaytt cttatcaatt cattcctttt ctcctgtggt ggtgctttct 126780
ttagtttctc atgccttcat tgaggaagct cctgacgcga ctgagtgcta gtctctagct 126840
gcagggacac cgtgtgcttt atgtggcatt acttacttgg gcttccacat cagttaactt 126900
ccgcgtttgc tccgctgttt ggttcaacag gtttgtccct atttctatca tcacagccgt 126960
ctggttctgt actgcattct gctgtatctc taccatttct ttcttcatgt tgtcctggat 127020
ataattctca agctagaaaa gaacagtgtt ggaaggcagt cattagtcaa atgaccggaa 127080
acctgattcc taaatgtttg tcatctcctc cctatcttta aaaaaaaaaa aaaaaaaaaa 127140
tctatcaaaa gacttgtacc ttgccttccc ttttggaatc ttactatttt tttttatcat 127200
taggaaaata cagtgtgatt ttatttttat gcaaaatctg gcaacttagt cacatcatgt 127260
aaaggaggga gacaagctac tggttgcttc tgtgttcttc tagaagtcca tgtcatggca 127320
ggccacagag ggtggtgagg gcagccacag ggactgctgg gtgctgccac tgtggggttg 127380
tgtctgtcct acccagctgc aactctgacc atgcagtcag gaaatgataa tttgacacaa 127440
agaagcatca ctatttctct cacattctag acttttggtt tctccacata gacttgagaa 127500
gacactctaa gacagcatat aaggagagga gcacccttt gattttcctt ttaacctacg 127560
gaatcaccac tcagttccac attctgtggg gtcttcccca ccttcctccg tattgagtta 127620
attcgaccta ttaaattttt cctaacatgt atgcattttt cacaattttg tcatttcatg 127680
tatcaagcaa acttttaatc gcaccttggt ccatttatca cctaacgtgc catgggctgg 127740
ttcttctctc cctcagttac taaagatgat gatcatgccg actaatttta gcattaactg 127800
aaacacaaga gaaggaagaa gctcatttca ctgccattgg tatagctatc cctgtctatg 127860
gcagtaaaat tacatgatta tgtataactg caagacaact gagtacgtgg gaagagcctt 127920
tgggcttgga gccagggaag cctgccctct gctttatagt cttggttcta ggaaagttgc 127980
ttaacctttt gggaccctag tttcctcata tgtaaaatag ggtttctggt tggtcagagg 128040
agtgtcttaa agaggggtta agctgtgctt ttaaagtcat tgtgtatgcg taactccaga 128100
tacttagcgt ttagtttctt ttttttttt ttttttttt taaataatct aatgatggga 128160
accattcttc cattccctgg tccaaagtat aagctcgtga gtgcacaaas catgttttct 128220
tccttttcac atagtgtaac aaacattgtt tattacattg aataattgaa agatgattat 128280
aaaactggtt ctggtgccct cctttaaaaa cttagaattc tttatagagr aamcattcgt 128340
ggagtcagtc atcagacatg atttccccca aaatgttaac cactaaataa ttctgtgctt 128400
tctgtcttta agagtaggaa aataggatgg gaagggtaga gtttctctct tagagcttct 128460
ttgttgatgc atttcataga ttgtgtcttg tgactggtat cagatggttt taggattagg 128520
ctggaactat aagtttcctg tttccgatgc cccctcgcca tcgactctgc cccacttctc 128580

taagctccca gctmcctgca tgcccctcag cctggtcact aaggctgcct ccctggcagt 128640
cgttctcccg tggatattgg atgggtcaga tgagcaggat gcatgasagg cacagtcagc 128700
cctccatctg tggcctccac atctacgggt tcaaccacac catcaatata tttwaaaaaa 128760
aaataacaat acaacaataa aaaacaaaaa ttgyaaaaca atacagtata gcaactattt 128820
acatggcatt gacattgtgt taggtattct aagcaacctc gagatgattt agagtatacg 128880
agaggatgtg tataggttat atgcagatct accctgtttt acggaagagg cttgagcacc 128940
gtggattttg gtattctcgg gaatcttgga atcagtcccc cacagacatc aagggacagt 129000
tgtactagag ctccaagcat gtgtaaaatc attttgttga aatgttactc aagccatcca 129060
cccgcctcag ccccccaaag tgctgggact acgagcgtga gccagcacat ctggctgaag 129120
ccttagtttt ccatatgaac caaaacagag tagaccacta ctttaaaaaa ttaaagtatt 129180
aaaaaatttt taaaaattta aaaaaataaa aaatcagtca ctgatacccg gcaggccagc 129240
aaccatctct attataggct tcataaaata tgaagagtct gaaatcttac taacccttc 129300
tcagagttag ctcaggcttt ttagtgtgtg tgatctttct taattcattc tttctccttc 129360
ctcccctgcc tttataaaac tgtaacttt gtgattgaaa taaactattt aaaagaagcc 129420
ataaatagca gttcgtaatt ctcccctccg ctcatcgcca tgggagtaat ggaatttttg 129480
aggttgcagt taaagctgtg tgtcacccag aggcactgtc ttagttactc ctcacagcac 129540
cccagccaag ataatattta aaaagtttca ttccgggagg cttggaacta tagagataga 129600
ctccagctgg agtttagttt aagcccatac tcagaaataa taatttacaa agtggtataa 129660
ataaaaagtc ttaacctcct tcttgatttc agtacttaag agctaaataa aaattattgt 129720
attttgtcac tctaaatcat acaaccagag agggaaaatg aatcctctaa tactgccttc 129780
ccccatttct agagctactg agtcagatgt gtttgcaact ctccagagat atgagaggat 129840
tgtttataat tgaaaactta aagtcaaatt ccaatttgaa attaaactta ggaactttga 129900
aagacataca ggcccaattt taaaaaataa aatttcttaa cctgccatat tgttttctaa 129960
acataaaaac aatagaatgc aagatccttt ttaaattgct acttttagc tattcaggat 130020
gactaagtat aggttcacag tgggtgagct aatgtgtgtc catttatgtt aatcttacat 130080
aaaagcagat tacaaataca catgatgtgt gtatatacag ataggtatat agcatatatg 130140
tatatagtgt ctataaatat atacagctct tgaagcatgt atcatttaaa taaaagaaaa 130200
ttctgtgtga tactgactgc attgctaatt aattgaagtc tttgggagaa gaatggaaca 130260
gaaccaaaaa tgtgcagtag tagatatttt gtgttgattt aaaaagatat ttgagccagt 130320
cgtgatggct catgcctgta atcccagcac tttgggaggc cgaggcagga ggattgcttg 130380
ctctcaggag tttgagacca ggctgagtaa catggtgaaa cccatctcta caaaaaatac 130440
aaaaaaaaaa aaattagctt ggcctggtag tgcgagcctg tagtctcagg tactggggag 130500
gctgaggtgg gaggactgct tgagcccagg agagcaaggc tgcagtgagc catgatcgtg 130560
ccactgcact gcagcttggg cgacagagtg agaccttgtc tcaaaaaaag aaaaaaatta 130620
aaattaaaag taaaaatact tatgttctta ctcttgaagt cattaaatta aggttttaag 130680
agaaatatat gatgtgacag tcaggtactc tttaaaaaca aggaagaata ctgtatattt 130740
agccccagaa acactagcga caggaacagc cacagtaatg gtaggtactg tttcttggtt 130800
gccgkcactg cctgtgctgt atgggaatcg ctgtgtcggg atcccaggcg cctcacatca 130860
gcacaggtgg atgcagggct gagcactgga atgaccctca gcaaaatgtt agctcaaccc 130920
agaggccgct tcatactttt ccagcctttt aagagccaaa agtgatatat ctcaaaattg 130980
gcttgagtat accttccaat tccaggcttc acaatgcctt aagaaaacag acagaccacc 131040
caccctcag tggagggcca tttttaccac cagaaaagcc cagaattaaa gatgaccaat 131100
gccaattcta tcttctggga gcatcctgac aaaagaatct gtgttttctt ccaaagatta 131160
gtagtaattt ttagagatac agaaagacta tggatgtcca tcatatagta taaaaatgaa 131220
catttccaaa taaagatgtc ccatttaatg tagcctttcc ataaatcacc acgtatcaag 131280
gataatgaga acaaacctag aaacaaagcc atctggctca tccacttgga tagacagacc 131340
ttgaaatttc cctgtctctt gaccttgatg aattagttat tttctagttt attgtcctag 131400
aatgtctttc tgtttagtgt ctctcttatt tttactggct gtgactgaaa cccagaaata 131460
tagaaacctg cccagaaata tgaaattcca ttctaagtat aaggaagtct tagtacaagg 131520
aaaaaaaaaa aaaaaccaac ccagtaaata agccatcctc cactggcagc accaaactcc 131580
acttgccttt ggagaatgtt tcccatccct gtcatctgca ccgaactgct ctcatcaaaa 131640
cagttccaag atacttgaac ctcccgtggg aggggacccg gctctttcca atttcacatg 131700
catagcatgt gaaacatatt catgtttcgc aggaatgttt gccatcgcct tcatatctga 131760
agaggattat tccatgagcg tgatctgtag gcacacgtgt ctgaataggt cctgctgtat 131820
atgtgtgcga ggacagtgtg tgttcatttt gtcctcttct tgatggttga cacagtcggc 131880
aaagtgtggg gccttgggct gttcttcctt tctcagaact caagtgagtt atgcaagttt 131940
aacattgagg gccacagtga tccttctagc tgcatggttt gctgcttagt gttatttgat 132000
ttgctaaaag agttgcgccc cagacatagt ctttaaaact tggcagcgca tcgaaactca 132060
agcaaccagg atgaaatatt ttaatgcaac atatatatat atatatgttt acattaatat 132120
atatatattt tagtgcaaaa tatgttctga agtttttat tactcccaca acgttttgaa 132180
tgatcaaatt tgacaggaaa aataggtcca tttgtgaggc aactatggca gattgattac 132240
acatttaaaa gtttatctgg ctatcttcct tctcaccaag attgtcatca ttatttttta 132300
taccaaaaga aaagtaatct tgaaactggc tcagtaaagg aaaacataga taatatatga 132360

gtagcatgca actcttcata aatttgtcac ttctttgaag gtccttgtta tgagttgaat 170220
tttgttctcc gaaaattcat gttgaagtcc taaatgccct cagcacgtga ccgtcttcgg 170280
aagtagggcc attgcagttg taatcagtta agatagggtc atactggagt ggagtgggcc 170340
cctaatctaa tgtgacagat gtctttataa gaggacggtc atgtgaagac agatacagga 170400
ggaacgcctc gtgacaacgc aggtagggac agggtgaagc ttctacaaaa cagggaacac 170460
caaagatgag cagccactgc cagcagttag cagagaggcg tgggacagat cctgcctcgt 170520
ggctttggct tccagaagga accaaccctg ccccacacct tcacctcaga tttctgctct 170580
ccagaactgc gagagagtgg atttctgttt aagcaagttt gtggtacttt gttacaacaa 170640
ccctagcaaa ctaatacagc ctaaaaaaaa aaaaaaaaaa aagtaatagg aaaggaatta 170700
aaatataacg ctaccttgca gcctccacca aacactgttg ccatttggtt cttctccttc 170760
ttgttcaacc tcaggagggg gtgaaaaaag tccaggcagc tcctggtgat agctatgcaa 170820
agcttcattc tgcagcagta aaagtgtttc ctagaagtac taaggctcgt taattgcagc 170880
caccctataa aagaaggtcc tctttcatga agagcctgtt tctctgcagg aagatggggc 170940
tgacctcagg gcctccagca cttaggcact tatccatatg tctgtaacca ttgttgtgag 171000
gttagttgat aatggctcat tatcctcgct aaaatgaact cgttgaagta tgaggccagg 171060
ccttattgga atccttccct ttccctttcc cttcccgttt ccttttccct ttcccttccc 171120
cttccccttc cccttccccc tccccgtccc tttagatgta gtctccctct gtcccccagg 171180
ctggagtgca atggtgcgat ctcagttcac tgcaatctcc acctcccggg tcaagcgatt 171240
cttctgcctc agccttctga gtagctggga ttacaggtgc ccgccaccat gctctgctaa 171300
tttttgtatt tttttttttt tttttttttt tttttttttt tttttttttt tttagtagag 171360
atgggttttc accatattgg tcagggtggt ctcgaactcc tgacctcagg tgatccgtcc 171420
gcaggtgagc cacccgcctc ggcctcctaa agtgctggga gaggcacagg cgtcagccac 171480
agtgcctggc ctactgtctt ctctaaaatg gcatctgtgc attcatctca gccgcccctg 171540
ctcagataaa agcaatggcg cctcctttga aatctgagag acgcagggcc ctgcccattc 171600
tgcggaattc cttctccctg ctgcctgctg tgaggaggcc ccctttgcca cggaacctga 171660
aattcctgcc actggaatta cgctctggac aagcggcaag atactccttt cagtcccagc 171720
cactgggttc ctgctgcaca ggaggccagg gtgctgtgaa cctgctctca gccccgggca 171780
aagggaatct cgttaatcca ggtggccagc gcctcttcct cagagcatct gcagtgctgc 171840
agacagggcc tccctgcgtg gggcttctgt cctccacact gtggtgctgc tgggatgttt 171900
tcatggggcc tttcccttcc cgtcaccacg tgtgctccag aacccggtgc atttggatga 171960
agccactaga tgtataggtc agcagctcca catagaatcg aattatcaaa tgcacactac 172020
ctgatccaga atagatcgtc ctggggtaaa cacattcaca tattctgaat gtacaaatgg 172080
ctgtctagta aacacactgg aacttccata attattgtcc ttccagataa tttttcaaga 172140
ttatatgcac gtattctgcc attccttttc aagacaactt tagaacttcc tttggacagc 172200
tactgtaagc caaagggctt gcatttgaat atcttgcatg aagctaaatc tttgttcatg 172260
aaaggcagaa taattttata tgccacaaag ctgcagtagt gtgttaggtt tagtagatgg 172320
ctaagcacta cactgtatta ttctaatcct attttcacaa tttaacaaat gtgagacacc 172380
gtgctacttg tacaagagat acaaattaag gaatcttcaa tgaccttgta gcctagaaag 172440
acctttagta attcttctta atctccctac agagctaagt gatccagagc tgaattaatc 172500
cagaatctat gtcttcctcc gcctccggag tagctctaga aaggtcaaac ccttccgaga 172560
tggagtgtct gtgggggtag gtcctctttg ctgtgtgcga tcctgtgaga cagcgggatg 172620
tcctgcatct ctgaatttga agcgaggagt ttttctgcta tgtttgggga gagcctcact 172680
cccctgctca gtagatcaga cgtgttctct tctttcacca cagctacaaa caacacactg 172740
gcattgtttc ccagacactc gactgtcccg atgggcattt ggacatggtc tatgagagga 172800
ataagctcca gccactgtag tggctcatgg gagagggaaa tgggtagaaa ttctttccca 172860
aactggtatt tctagtaaag cactcagcca gagcctgcag ctgttcacta ttccatatca 172920
attctaaaca gcattttcgt tggcaaaaga aaagtgagaa aacaacaaag cttgaagccy 172980
aaaactttgg gaaacccctt tcctgaatgt gtttacttag ggcttaaaaa tatgcctgtt 173040
ttcagaacag aagaactaat atccatgttt tctatgccga tttttcagag tacattttaa 173100
atgtaagtac atttagtgat taaaagggaa aaatacttga tcgttttcta aacataacca 173160
aaatctcact atgtaattgt tttttcctct atttaagagc agaatatttc attgctacca 173220
aaatgctagt attttggaga aaatagaaga actagaataa gtagtcagca atacaaaacc 173280
ctgcgtggaa gatgtgtatt ttggataggt gtcaacatgt ccaagctctc agtgacaaac 173340
acaggctcat tacaggtctg agcaaatgtg ccacttctca ggaagacaag gcagatcaat 173400
gtaaaggcag gtggcacctg gtatggctca gactcgcacg tggttctcca cagagctgct 173460
ctcggctcct ttggaagagg ttcaacgttg ggagcacagg ttgcttctct ggcccatgtt 173520
attcctggag ctactacttc ccagggcaga gttcgtgttt ttcgttcata aatggcctgg 173580
aaatcctagc attgggccag ccatccagaa cagtggagct gcatgatctg gtctggggat 173640
atttcaaagg gaatagaata ctgaggccct gtgggatgga ggctgcttcc cgatattgag 173700
aactgcacca gactgagctg tgtccagagg aagggagaac gtctttcatt cacttaaaac 173760
tcacccaaca cctgacacct ccatcttggc atcatccacc tgtagcctct agccctcttc 173820
atctgttaag tgagagtaac tggcaggtta tttggagagt gaagtgacat cggcagagtt 173880
ccaggtatgg tgtctgatgc gtgagttcgc ccccttccc ggtccccttc tcctccattt 173940

aaactatccc caacttggag attctgatgt tgatttctca ccaactgtag atgctggttg 132420
agagatcctt tctatttaaa taaaattcaa ggtccttaga ccttttttact aattagtttt 132480
tgtccatctg agtgacctaa ggtggacaaa aactaattaa tcagagggtc taaagacctt 132540
gaattttctg gtaaattaac aaataatttg agatttcctt ggaaactttt tactgttgcc 132600
catttcaatt tcgaaatagg attttgcaac catctctcac acacacatac acgtttttct 132660
atcccaatga tccatccatc ttcccaccca gcccttccat ttttctagta aacccttgaa 132720
tttttctagt aaattcacaa ataatttgag atttccttgg aaacatttta ctgttgccca 132780
cttctattta gaaataagat attgcaacca tctgtcttac acacacacat acacgttttc 132840
taccccccat gatcccatct tcccgcccag ggtccactgt cctctctgtc tcttgctggc 132900
cacgcatccc ccagctgctc tcctttcatc ccgctgtcag agtcaggaat ccacatgcaa 132960
agctggtgac cgcagctcac tttcttccct tgcaggtttg cctgaggata aggtccagat 133020
tccttctttt taagacacac accgcctcct gactggcgcc cctgatcttg tgggcctcag 133080
acctgggcgc gcactgtcct tggctgtccc agctgcccag cggcctcata ccacggccgc 133140
ctttgtatct cctccttgag tcttctcttc ctcctcaggt cccaccatcc cccattgcat 133200
gccctmagca aagatactcg ttttgtgttt ccttttgata tcaaaaccat tttgtatttg 133260
tgtcatttca ttttaatctc cacagacaat aggttaatgt tcttgcttgc ttggtgaaga 133320
gtgaacagaa tcctcaaact ctgcaaccat tctacatata caccctagta acaacaagca 133380
aaacatccac tcttagaatt agtttgaaaa cttgagtgta agattattaa atccagggat 133440
attctatttg ggaggctttt gacctaatgt tcttggttcc ctgtcatgag gaaactctga 133500
aacatcattt gaggtctcca gacagaaaag tggcaaaact gggctctcct ccccctcctt 133560
ttagagttgg gcttgtgtgt gtgtgtgtgt gtgtttattc tggagatttt gctgcctaag 133620
cagctgtgta ctcagcagta cttcatggca gaggctgagc ctaaagaggg aagggctggg 133680
agatgcggat tttgggcagc actttgtcct cctaaacccc tcgccagagc ctgggggtta 133740
ggcacagtac ccacagtgag aggtgatgtt cacatgccct gtgacgtggg aagcaagttt 133800
tctccatata ttgatgccag atttgaattt ctagaaccta gaaaagccca tgccaaagct 133860
acttgccatc tgttgactgt ttttatagtc ttggcctttt cttcacgttc agtgtaaggc 133920
cctagaagtt gaggcaaaag ctaaaggccg agggagggaa gcctggcctc tggtgccaat 133980
ttcctagtgg gtattgtgac ttctcttagg gagcacactt gccttcacct gccctgacca 134040
catggacgcc tgcccacata gggtctttta agcacttcct gaaatggatc tgttctgatc 134100
tagccttttt gctttttttct agtcatactt ttttattgtc tttttttga gatggagtct 134160
cgctctgtct ctcaggctgg agtgcagcgg cgtgatcttg gctcactgca acctctgcct 134220
ccctggttca aacgattctc cccgcctcag cttcccaagt agctggggtt acaggcgcac 134280
accaccatgc ctgggaaatt tttgtatttt taatagagat gggttcgcca tgttggtcag 134340
gctggtctga aactcctgac atcaggccat ctgcctgcct tggcctccca aagtgctagg 134400
attacaggtg tgagccactg tgcctggcca tttaataatt tatgagtgac tatctgatac 134460
tgtatctaga taaccaaccc ctttcctact ttcgctagta taagagactg aaagttcact 134520
tttggccact atataactcc aagatgtatt aggaaataag tttgtgggcc tcagctggtg 134580
gcattctaac attaatagtc catgcctctc ctcctgtgga taggtacacc ctacagtaat 134640
ttgagtgtac cagaatgtct gtgctctggc aaatcctatc cgctttgctc ttctttgagt 134700
gcagctgcat attctttgca ttaatttttt tcacatatat ttgaatatat gtttttccac 134760
atatattcat atcattttac ctctttgtgt gtttccctta ccactactcc aaaatttgat 134820
aaggaaatgt gcttttccct tcaaaatgtt ccatttattt tctactgata aagtggctat 134880
ttctcatcaa tagcaggcat tttaaatata tgtaagttta aggagactgc tgtagtaacc 134940
tcatgtaaat ttctttgggc atttcatatg caaaaggtgt cacattttac acgagtgtct 135000
tttagaggtc ttgtagggca catgtatatt taccagatgt ctgtgagcgt gcagcctcat 135060
ggcacgttat gcatacctga cacttgcaca gattcctgga agatgaggag caaatacagt 135120
gcaacagacg ttgtcaggcc acgtctgcat atatagatat atacacagca agaatagtta 135180
cagcagctta caatgacaaa atgcttctca gtgtgtatgt gtgtgtacct ctgtctcacc 135240
agattctcac actgccttag cttgggtttc cccaaaagca gagcctgaga caaaggcagg 135300
catgcaggaa gtttatttag gcagtggtcc cagagcgcag ccatgccgaa caggcgcggg 135360
aggcaggggc gccgcagggt ggttcrcaca cgtggactca ggcggccacc gccgcgctgg 135420
ctggtaagaa gccccacagg atctcccaag gagccctggg acagtgtctc agaacatcca 135480
cctggggcaa gaatggggac tgctgtcccc agggggcagg tggagcctag tgggcattca 135540
tgacccaggt tttggagctg tgcttgcgag agtgccgagg aggctctcat gggtgtcccg 135600
aggcagcttg gagccaacgt ccctaggcat ggcctggggg tttgtgggaa ggcctgaggc 135660
aaggcctgtc tctgagatgt cctgaagagc aagttgggcc cagagggtta attccgagca 135720
gcacaagagg gtgaattctg agcagcacca gagggtttcc ctgacacagc aggggatgct 135780
ttgaggcccc tttaatgaag gagaaaaatg aggcttagag aaagtcagtg cccaccccaa 135840
gtctcatggg ccccaggctg tgggcagtgg ctaaagacag gctagtgggt aactcggggc 135900
cacgtggaag gggagcttgt atttatagcc cccagtcagc agcgctggag aggagaggag 135960
aggaaaagca gtgctctgag aaagacaata tttctagtag attggggcag ggcaggcctg 136020
gagacaggaa accaaagcca gggttgtcat gcaggagtga gatgaggttg cagcagcaga 136080
gcgaatgcgg agacgctcgg caggtccttg gtggcctctg agttattctg cagacttctg 136140

ccattcgtct attttttggg atactttgtt aaattctcag cttagaagat agtagtgatg 136200
ttttccctac agagatagaa gaaaatataa aactattttc tttttaaaac tgtactgaat 136260
gtagggccgg gcatagtggc tcactcctgt aatcccaata ctttaggggg ccaaggtagg 136320
aggatcactt gaggtcagca gttcaagacc agcctaggca acatggcgag actccatctc 136380
tacagaaaat ttttttaaaa attagctgga catggtggct catgcctgtg gtcctagcta 136440
ctcaggaggc taaggtggga ggattgcttg agcccaggag gttgaggctg cagtgagccg 136500
tgatcgtgcc actgcactcc agcctgggtg acagaatgag actcaatctc aaaaaaaaaa 136560
aaaaaaaagt actgaatgat aaatgattac aaatagaagc agttaaaatt tagctctagg 136620
aatgagattg atgcatttag gctacaatat accaggaact tccttttttaa atgaaactag 136680
agcgttcttg cctttctgaa tttaaggcac actgaaagaa aaaataataa taatgtaaca 136740
aaatgtctca gtgtttttct atgccaaata gaatcttatg tatatctgtc tagagacata 136800
tatgcataca tttgtctaca catgtgaggt aggggtgtgt gtgtgtgtgt gtgtatctgt 136860
gtgtgtgtgt atgtatgtgt gtgtgtttca gttctctaag aaacagacat tccaaaactt 136920
gtgtgtgtgt gtttcagttc tctaagaaat agacattcca aagcttggtg ggcaatggcg 136980
ggaggcttta gaccctgtaa tattgtcgga gtgtcactgt aagagggacg ctagcgcctg 137040
ggtcacagtg ctagctggta gcagagtact aacttaaacc ctggtctccc aacaccccat 137100
ccagcactct catcgctgta ctgatatgcc tattcttatc ttaaaaaaaa aaaagtgctg 137160
tctgaggagc attaacattc ttacttttttc atttttgaaa tgaagtataa agatactgat 137220
ggccttttac gtctcctctc tgccctgttt ttgctgtctc tttctgtgtt acatgggttt 137280
gccaaaatac tgggtggagc tctgtggagg aggtagcatg atcatctctg aagtgggcag 137340
tttttttctt tttccaataa actgaattta cttggtcaca atgactatcc taaatggcta 137400
gaaaagggaa aaggctagcg aaacttagat gattttctaa atttagataa ttttctagaa 137460
gacattttca aggcaaacta gtttttgctg tcctttataa ggccggcagg aagcgtgtgt 137520
tgtttctgtt ttaaaaaggg agaggagcgg acttgggaat gctgatggga atgcttgaga 137580
aatctcacag cagggctgtg cgtgccctgc cgggtcccac tgcctctgga cagaaacccc 137640
cgcaactcca cccccagcca agactttctg cttctttatc tcctctttct gctagcaccc 137700
aaaaagttga aagaattcca atggatagaa tttttgagat aatattggaa gatgctcaaa 137760
atacacagga ttaatttaca cgaagactca gcgggaacac aagccatctt ctgtacatga 137820
agatgcacta ctgacccgcc gtccgcaaat gtgtttgtac agttactttc tcagtatggg 137880
tgatggctct ccaacgaact gctcctcgtc tcctgcctgg acacccttct ctctgtgctt 137940
tcctgggtta gagtaaatgg atgcaaacac acatttccgt gctctgcagc aacttgagac 138000
tcctgtgagc aaaacgcact gacgggcaat gtgcgtgggt cgtggggagc atccagctcc 138060
catctgcgga ataaacccgc ccaaaccata gggaaaagcg ctgtcgtata aggccagggg 138120
attttcagaa aagaaatgtg ttctttcctc tttgattttt gtgttcataa agctgtaggt 138180
gcagcttttt ttaatgtaat gatttcataa ccgctgaagt tcgtgctttt ctgaactatt 138240
taggaagata atctacccct tgtattggat gagatgatct gtcccttcga cctctggttc 138300
agttcccatt ctcccaagta tttaaagctg cgagtttttt catattttca tatttattta 138360
cataatttaa acccccctgtg tgcatggact ttaaggagct gtacatctgc ctgggctttg 138420
cagaagctga aagggcgcaa tctttttata actcacatta gaaacacaga ttatttaacg 138480
gggctatgtt ttgcacctta atctttaaag ttgcaatata ttttaagcat tttaaccttg 138540
ttttagatct gatcagcagt agaatgtttt cagataagaa acaatggagc aaaagcaaaa 138600
caatattcaa tacctagatg atgtggcaag acagagaata gtataacttt ttgttttcca 138660
aatataactt ttatcttcat ctcttgatct gaaatttggt aggaagtgta acaagtacga 138720
atcaacatat ttaccatttg ccatttcaaa tgttgatagt gaagctggga cctctgttta 138780
ttatggaaga ccatagaaaa ccccataaac acgttctact tctgtctgtg gccagcagtc 138840
cagcaaaaat gttctaaaag cacatgcact gtgttccgtg atgattatag tttgactgtg 138900
ctggaaagag agactgtgaa ctgcacatgg tgattatgac tttgggcaaa tcactgaact 138960
tgtaattatg tcttccaaga cctcctaacc caaaataaga gagtatttta ctacaaaata 139020
tatgttacgt caaactgttt tttacaaaat accagctcta gggatgtttc caagtcattt 139080
tcggagagag tttgtcgaag tttttttcag ggtgtgtcat tcatgtattg gaggggggaga 139140
gggttgagta agaaccgaca tgcacaactt ggccatgaaa tgaagcgcaa gcacatattt 139200
tatttctata ggattcctca ttctaaagta atttttacag aaaatggcac tctaagaagg 139260
aattcattaa gataaagaca cagatacagc atttagagtt acactttgcc ataaaagagc 139320
ctcccttacc tcctgacttg aatctataac atctgctgaa ctgtcgacat caggaagact 139380
cgaaatatrt tttgaggcca attatgtcat ttcagattga acctgctaac atcagattct 139440
ttggtcagta gtctcactag ttttgttctc acaatggaat tattattttg atttttaaat 139500
gttgctccat ggagactggt atgatgagct catgctctgc agttccattt taacaaataa 139560
cataatagat cgctgtcaaa tgaatgccat cacagacatc atgttgggtc acagaagcca 139620
caccctagga gtgcttacta gatgatccat ttccatgaag ctcaagaatg gaccgaactt 139680
actaaaggtg atagaagtca gaataatgtc ttccgggtgg gaggggtttg aggctgtgaa 139740
ctgagcagtg gcacgaggga gccttctgga gtgctgaaaa tgtttcctag atcttgacct 139800
cggtgctggt tacatgagtg aatccatatt tttaaaagtt atcaagctgt aaatttcagg 139860
ttagtatact ttatccattt cctgtgtttt atatctcaaa aattctttta aaaactaaaa 139920

gacatttaga aatgaaatgt ttgacaagtt ttgttgtgac actatgaccc tagttactat 139980
gggtggtttt atttgtctct gcagttttca tctgggagca cctaaatcat gcctaatgaa 140040
atgaactttg gaaaagtaat tttaaagtaa ctatcttaga gaactgtgga ttaaaccatt 140100
ccagccatct gatgagggtt aaaatgtata ttcgtaatct gacattccaa aacacgattc 140160
tttccggatc aagcataaaa ggcatttgct cttggaagac caagaaagaa ttcatgtggt 140220
tcccattagt ctaaaaataa ataataaata aataaatgtc tgagtcatgt attggatttt 140280
gttggatttc agtggcttca agtataggaa gaaaatgatt tgtgctatta ataatagttc 140340
tacccattgc ccattggaat aaatacaaga ttactctgag aaaagtgaaa tcgattgaat 140400
ttagttctgc tttgagctac tgcaatgcaa gtgtttctga cttttgagac atagtataaa 140460
aaactgaata aaataacttt gttcatattg aattgagttg gggaagtagc gatatttgtg 140520
acattggaag ccattgtcat tacagattca tcattacagt acagatttga gaatcaaaca 140580
cacccggtgt gagtcccagt tcagttcctt aggaactact ggctcaatac tttctaacat 140640
cacatttctg gttggtaaaa caggtatata aatacctact gtgctatgct agtgtgaaaa 140700
ttaagtggaa tgagttagca caattcccaa actgtgtgcc aaggcaccct gggtccttgc 140760
agcaaacaaa ttggagtaag ggagacagtc tgaatattca agggcaaccc agcagtgttc 140820
aatgactgtc agccactgga agaattcata gctcaaagca gctcaccgtt tcaacagtat 140880
cagcttgtac ctctgtaaag ctaggttttt ggtggttgct gtttgtgaaa agcaagtgct 140940
acaggaaaat tagcatagag cgtaaaatgc aggtggcagt gtccaatttg attccaaagt 141000
ttgagaagcc aggtgtgcct aactggcaaa tatattccat tcatacgtca ctggggttac 141060
ttaagaaaga aaataaagga tctttttttt aaaactcaat ttatatgtat attggtattt 141120
tcaaatagct actaaattgt cagaacataa atacttaaat tatttggagc taaccgctta 141180
atacaagcaa ctgttggcct agagataaat agagaaaaaa tagtgaatca ctaagggtcc 141240
catgagctga gaaagtttga gaacaactag cctagaacct tgcacttggt aggatataaa 141300
ctcaacgtac cctcttcctc cccttccccc aatccaggta ttgcctttaa ttgtaatctc 141360
tatgatttga tatgtttatc ggaaagcgag taagtcaaaa agaactaata aattgtgtaa 141420
gaccttcatt aagatgtacc cttccgtgtt ttcctaactt ctgaaatcac taggaaaaac 141480
agccatgttc cttgcaagct ggctggttag tcctgtcttc tttcaggtga acagcattta 141540
taaccacggt gtacctcgga agaagcgttc tcagagcaac atgcacgtgt tccgtgtgta 141600
ccgtggttgc cttcgggcta atgcgttttt ggaagtgtag attggtgcca gttttacaaa 141660
actcatgtgg cctatttctg cttgtaattt atagtttgcc tcctcaccat cctcacttgc 141720
tctaaggtga actagtttta taccattaga ttatacagaa aagccaaatt tacttgcatg 141780
ccacagcaat tcgggaagta aactttcagt gtgattctcc aaatgcttgt ggtaaaagta 141840
cagagacttg aatcattttc ccataattag tttcagttct ggaggcccgc cccctctctt 141900
taaccctctt gccttgcata tgtgtctctt gcaatggaag ctagtggaaa tttcctctcc 141960
cctttcactc cactgccata agttataaaa agcatgccat tcagaactga cgttttcttc 142020
tgcatgcttg aatttttact caacaactgg aaggggaaga agttatttcc agagatgttt 142080
ctctgtttat caaaggggcg cagagtcaca gtagcacttt ggaccaccgt agaatggctg 142140
actcacttgc ctcatcagta gaggggcaca tgttcctatc aaacagtgag tgcctttgag 142200
tgacggtgtg tgacacacag cacagcagca ccatttctca gagctgcagc aacactggtt 142260
cacacaagtc actagagatt cccatctcca ctgactcact cggtgggaac aaaagctccc 142320
atgccggtgg gaatgcgggg ccaggggagc accggaagaa gggagccgtg gcagaggttt 142380
tcctattgtt aggtttgttt gtttgtttca gtgagtcatc ttacccccat ttttttttt 142440
ttttttaacc aaaactcact gtggttactt ctctagtttg ggttatgact cctacgagcc 142500
agtttaattt tatcagtggc agtgaattct tgaacgcttc cctcagttgt agaaatttag 142560
tttcacattt aagtggtcca agtgccagct taaactttgt ggtttagtgt ctttactgaa 142620
tccccccctag tggaagaacc tacattggag ttttggctgc tctttgggat tcaaattatg 142680
agtagttggt tccactggaa tcttggcctt ctcctggagt ggcttgaggc ggcacacttt 142740
gactttagaa ggccaaagtt agaacctacg tgaaggcttt gtccaaaatg cctttcctgc 142800
accctggcat tttcaggtgg tgtgtgtaga cctgacagga ctctactcgt gcgtcacttc 142860
ccagctgttg gtcccctgcc acttagatgc ctttcatggg caagtccatg ctagcctagg 142920
aaatttcccg aaaatggcag tgataactca gaattaggat tttgatcctc atctcaacca 142980
ttatccccag tggtcgctag ctctcctctg cccagctcga ggaaaatgct gggtttttca 143040
ttgcagctta cttgcatttc agtggtccca gcaagcactc aagaggaaag atcagaccag 143100
ccaaacttgc aaggcaggct gatgagccaa ggtacagaaa gtacacctga tggatttttg 143160
taatgatggc cgtaagtcat agaaggtgaa gcttaatgca atttagaaag atttgaaaaa 143220
aggaagaaaa gtcctcgtgc tcacagaagc aagctcccat tggcaaaatt attgtgtata 143280
acaagcatct ctctctgatg atgctggaaa aaaagaaggt gctatcaggg agggagggaa 143340
gatttgaaag accagagtga gcagcagata ggcccttggg tccttccttt tattcccacc 143400
ctcttttcca ttggctgttg ataaagtttc atcacttttt agctgcctgt tgtatttcac 143460
tacctccttg ttaacctctg gttataacct gggggggaatt atgtttaacc agtacttaat 143520
gaatcaatct attcctcaaa agttggttct gggcatcacg gaataaacac caagaccact 143580
cagcgtattt tctgaagagt ggatttcatt agaaggcagg ttagtcttgg aactcagtca 143640
aaacagcttt cagtcagtcc aactccacag atttcagaga tagcagtaca agaaaaatga 143700

tacatgggtt tgtacagttg agtgttgaag tgttgacttc cataaaagaa cacaaatgtc 143760
aatctacagc agtaccatgg gtacgtaaat tgtgttatac agcgaaacac tgcacggtga 143820
tggaaaagaa gaaactttaa catgcacaac aacatggatc catcacagaa ataataaaag 143880
agaccaaaaa ggagtatgta ctgattgttc caattacatg acattcaaaa cctagcaaaa 143940
ttaacccatg gtgggtgaca gagcagagtc aggattggcg ttgagaagag ggggatattg 144000
acgaggaggg gcctgaggaa gccacgtgca gggtgggaag ccttctgtat cttgagctgg 144060
gcagtcatta cacaggtgcg cacatatatg gatgaggagg ggcgtgaggg agccgcgtgc 144120
agggcgggaa gccttctgca tcttgagctg ggcggtcgtt acacaggtgc gcacagatac 144180
agacaaagag gggcgtgagg gagctgcctg caggggaaga agccttccgt atcgagctgg 144240
gcggtcatta cacaggtgcg cacagatatg gatgaggagg ggcgtgaggg agccacgtgc 144300
agggcagcaa gtcctctata tcttgagctg ggtggtcatt gcacaggtgc gcacagatac 144360
aaaaatttac caagttgtac actcaagatt tgtgaatttt attctgtgta agttatatct 144420
aaaaaaagaa aagaaaagaa aaagctagat tccctaaaac agagacagca gggctcgagt 144480
ctgagctagc tatagccatg ccagcagcta gatccatgaa aaggttgggg ttggctttgc 144540
ccaggtgatc attcggggac gggggacgtg ctgtgaatgg aagatgtgcc tgctgtcagc 144600
actgatgttg cccacccttt atttctacaa cgctgtcttc aaaagaatta catttcaatt 144660
ttataccaac tatcgtgcct cctcatgaat cccttccccg cacaacctgg aaaccctcgc 144720
ctggcgtcgg ctccatctcc agatgttact cactggctac cgctaggtgg ctgcgaaggg 144780
tggcggcgtc actgatgcgc actcaggcag cagccatggg gaggttgaat ccccggggca 144840
tctgcctctc cctatgtgtg tgggtcctgg gagtgaggca gtgtggcgtg gggctgttgc 144900
acacaccccc gactgtaggg ctgcacccag acacgtgcgg tgaccccgtc tctacagccg 144960
cttgttgccc tggcaccaag ccaaccactc agcatccagc gcgtcctcac cctccctccg 145020
gggtgaagcg gaaacaaggg tatgtgccaa aactggcctg ctcaccattt cccagatttt 145080
ccacatttgt tcccactcgg ggtgaggggt gtgcttctgg tgtgacagct gtgggctgtg 145140
tagggtggcg ggcgttggtg gtgaagtctg tcggccctcc tgacccacac acgagggggt 145200
gtggatttta tattgaaatc tttttaaaat ctgttttttt gtaagaggct ctgaaaggaa 145260
gaaattttat cagagttttg cggcctgtgt acgttctgat acctctcaga gctggagttt 145320
cttacccata taggacaagc tgttgtgaaa ttgagtgaga cgatgtaagc acatggcgtg 145380
cacctgataa atgccagctg ccaccacagt gatggtcagc agcgtggtca ccactgtcgt 145440
ttcacaatta cagcccaagg agcccaaggg gaaggagtgc ctctctctgt tttgaccttc 145500
tctgactgct gtcctaataa acagtgtcct ttctacaaga accctgtaga cttttgaaac 145560
caacaagtga aggcactcca aggcccttgt tttgagaagg ggtaagtgtg ctaggtaagg 145620
gatttccttg ggtgcttacc ttccacggct cctgggcccc tgactcgaag ctgaccatct 145680
gtgctgatgc tgacttagga ttttaaatca cttaaatttg agctggatag agaaagggtc 145740
ctagttaagc tgagagggct gcttattcgt gatttttttt ttcttctttc tcatgcagag 145800
actgtttatt ttagtggtag cggtatttag gggtgaagaa ggggaaagga agaatagtgt 145860
gccatcaatt aattctatgc atgtcagctg caacgccttc atggcacggg acaggccaat 145920
tatgtaactg taaacaaatt atatgtatta aaagttgtcc aattaaagga aaaaacatgc 145980
atggatttat gtgtttgtta ttacccagaa gggagccatg ctgtacttga aaatatgcaa 146040
aatttcacat cacaaaatca ccagttgttg tttgaggggc tggtgttctg attagtctta 146100
attttttta actcataaca tttttgtccc agtcatcaac actgttaaga acatgtcact 146160
ggtgcagtta agttaaaaat gattcaggtc aggaattcct gtcattaaca atttttata 146220
ttaaagttgg aaaagtttaa ggaaatttaa gaacctattc cttaatagtt aaaaatagta 146280
aggaatttca tatacccca aatattaagc ataggtaatt agcttgtggt tgggatttga 146340
tggttttctg tttttcagca aaatacaata acgtactttc tcgagcagaa tttttacacc 146400
aacatttccc attaagacca gtttgtttag ggaattttta agctacatct gtatgtaata 146460
atttttgag attccaaaga ctacgcagtc taataaaact ctaatacttc aactatcttc 146520
agactaatgt ttataattac ccggtagatg accaagaatt gatatcatct gttgattcca 146580
gaaattatgg cagagaaaat gctgtcagga acccaaagaa aatcagagga aatggtacct 146640
ctaagaaatt ctgaatcttt tctactaaga tatgtggctt gactgcttaa ccccaaaatg 146700
cctgcttaga aggtagtttg gggctatctt gtaatactca tttagttcct gccttcttct 146760
gccatagaaa caacatgcag aagcagcatt gcttacgact cacactgaac ctgaagggat 146820
gaaattacat atgacgatgg aatgtggcca tattcacgca gtcacagcag tgtgttgccc 146880
aatgacagta ctggagcagt ttccacagag gcactcatgc aatatgcaga atacagacat 146940
tttacacaca cacttacgat ggtccttttc attgtcgaaa aggaattcat tatctttcga 147000
gtaaacatgt gctttgaggt atataactct gaggtataga agttagaaca tttaacccga 147060
ttagggtgac tggaattata acctttaact aatgtgagat atagtataga tcttgataag 147120
tgtctttctg gtgttcctat taaaattcat tataattacc gttcctgcaa ttgtgtagca 147180
tcttacagtt tccaacaccc tgtgctagcc atcatcttat ttgaaacaca taataaccct 147240
acaagttcac tgatgtaggt aagaaaactg gtaccgtttc tgaagataca cagtgattgt 147300
ttcggccagt taattaaggc aagagatcac tcaacaattg ttctacagtt attcctgctt 147360
ttttttttt aactcactca ttaagtgaaa gaagccagtc tgaaaaggct gtatatttta 147420
tgattccaac tgtatgacat tctggaaaag ggaaaactga agatagtaaa aggatcaggg 147480

tttgccaggg gttaagggga agaagggctg tgcaggtaga gcacagagga ttcttagggc 147540
agtgaagctg ctctgtgtga tgctacaatg gtggatccat ggcttcatac attggtccaa 147600
acccacagaa tgtacagcac cagatgtcaa ctgcgggctc tgggtgataa tgatgggtca 147660
atgtagattc atcagttgta accagtgcac cactctggtg caggatgttg atcgtagggg 147720
aggtggctgt gtgtgtgcca cggaggggg atatgggaac tctctgcact ttactctcga 147780
tgtggctgtg aacctaaaac tgctctaaaa aacatagtct tttaaaaaat catttactac 147840
atatgaaaag gaacaagtaa agcaacaaca acaaaatgtt attgtgtact ttcagattgc 147900
accagtaaac ctagccagcc ctcactaggg tcttctgatg gttacatagt taaaagtaca 147960
ctagcacacc gggagaataa cttcagaggc ttgctggtct aatggtaatt gcgtcggctt 148020
cacacgtcaa catttttta aaaattagat tttcttgaat ctgatcatgt ccaagatacc 148080
tcttattttg gtatagaacg cctttattca aacaacggga gaacatgaac atatcccttt 148140
gccatagttt ggctaaattc ctgaggctgg ctggggccag aaacaaaatc cctgaaatgg 148200
tctcaaaatt ttttttttt tttttacctc tccccttttc cttctggttg gtggtctttg 148260
gggcctacga ggccctcagg cagaggggaa atggcagttt ccccatcccc ttttgggact 148320
tcttgagcag aaaagcgaat gtcagacggt ccttataaag tcccacgtga ttcagccact 148380
gaggatggca ctggctgtgg atttacatgt aagacaactt catggcgtat tttcgccttt 148440
tgctgttgaa tataactacc aagatatggt ttgggcagac aaaatagaaa tcttctgtgt 148500
gtagcatgtc cagttggata ctgttagtga catagagaga cgagcgcaca actcaggttt 148560
aaccttcatc cctgaaattt gccggaacag tcataatgaa ggtgctaatg tatttcctga 148620
aatactgagt acttcagaca gggagatatg ggtggtatct agtagccttg tgataagacc 148680
catattagac taatagtagt cttatcacca gattaaacca cctggatagc ccacctcaag 148740
tcatcaagag tgttaacatg ggagtaagtg tgacaaatgc ccaggtggtc tggactaaat 148800
gtgacaaaat tgagaaatag accctacaag atctggattt taaaaagaga gaaaaaaaaa 148860
aatggaaagg ctggctgctt gcttcctttt aagactttgt tcacgttctc gcccccaaaa 148920
gccaattatg attataattt atcagcccac aggaaatgat tgcttctcta tgagacatcg 148980
tcaacatgat aaaataatcc atttcccaag atttctatat cttagtatct catctcttta 149040
aaaagctcca ttgtccataa aaaattataa aattacatat ttttacatga caggtaattt 149100
ttaatgtata tttttaattt ggttgttggt ttttaaaata gtaaaatatt aaatatcaac 149160
tatgaatatt ttgtggtggt aagttgtcag gttaatgtaa agattccaaa aataattcac 149220
agacatgtgg aaagttgctc agagggagaa ccagtctgat tttggagaaa gtaattacca 149280
tcagagcagc cctcggaggg agcgggagag tccacaggtt tcaatcaggt tctagatgaa 149340
ttgcaaagag aaaggtttta gctggttgca ggaggggctc tggtaaaagg attaagtcca 149400
gttctcagga gttttttaat aggtttcaca tcttttgtca actggtgcaa ggaaggatta 149460
ggacagaaaa gaaaggtgat ttcatggaga aatatctaat taaaatatta aagatagtcg 149520
gatggcacac ctgacctaga gtccaggcag tggtaggcag agttccttcc cctttttttt 149580
aaaccacaca taaaacagtc attttaattc caacaaatgg ttcatactgg tattctaaac 149640
cactactcat gatttttttt actcttttta tttacatcaa atcattcaac ttcacatcat 149700
tttcttttta agcattaaca taatccaagt gccaggccat ttttggtgat ccaatctgta 149760
gaatgtgaga tggacaataa caatcaaacc gttttcaaac tctaatagtg ggaagagaag 149820
gccacatgga acttccctga ggctgaattt cgtcgtcctg cctttcaagt ggtgtcctgt 149880
gaaatccagc gtttccccct gtcaacttcc agaacagggc tgtaactaga tgtatggttt 149940
gtaagaatat cccatgtata cttcctcttg gttataacat aatttgtttt gcgggggtg 150000
gtttgccctt ttttttttt ggagacagga tctcactgcg tagcccaggc tggagtacca 150060
tggtgccatc ctggctcact gcagcctgtg cctcctgggt tcaaatgatc cacccacctc 150120
atcctcctga gtagctgaga ctacaggcat gtaccaccac gcctgggtga tttttatatt 150180
ttttggtaga gacggggttt catcgtgttg gccaggctgg ttttgaactc ctgagctcaa 150240
gcgacccacc cgcttcggcc tccgaaagtg ctgcgattac aggcatgagc cactgcaccc 150300
agccacataa atttgttttt agtcttctga acgattaaat agttgtacca attataccaa 150360
ttgcaccaat tctattacaa ggtggaattt cttatcgttc ctttacaaac aggatattcc 150420
cagttgcttg tttttgcttg ttttcctagc agcttcagca ccatcctcac atagaagggc 150480
tggcatctca cctatctaga ggtgagaaca aagctgtgct ctcagcaatc ggaatctgtc 150540
aagtctgctg tggggacttg gtatctcagg cctgatgctg gcctaggagt gccctgcact 150600
cgtctcaaga tcgatgtccc agtgggcgag aattgctgcc aagactaacc aagggtgtca 150660
accagtgact taacttctca ggctcacttt ttttttatt tttaataaaa acaaattgtt 150720
aaagaggtaa tttaaaatat gtactatata ataagtacta cagcatatac agtgtttaca 150780
tacatatagg cattaaatat taagaatgtt tatttcagaa tcatataatt atacctgata 150840
tttacttttt gtcattcttt gtatattctc catttttgc agtatctata tattacttgg 150900
gtaataggaa tagcaaccat tgagaaatag ttctaattga ttttcctttt ataaaagggt 150960
ttccgtgtag tgaccaagga cttaacatca tccccacccc acagtccctc acacgcctga 151020
ctccctttgt gctgtgttta atttctcatt tcattcattt acccttctgt gcagcacata 151080
ctagctgctg ctacactaca cgttctgacc aaagcatagt gtccccctgg ggcaagactc 151140
ttgggaattg tctttttttа ttttttttтc atttttagg gtctcacttt gttgcccagg 151200
ctggagcgca atggcaccat catagttcac tgcagccttg acctcttggg ctcaggcaat 151260

cctcccacct cagcctccca agtagctggg accacagctg cgtgccattg tagagatggg 151320
ggtctcactc tgttgaccag gctggtcttg aactcctggc ctcaaagtgt cttcccatct 151380
tggcctctca aagctctagg attacaggtg tgaggcactg tgcctggctt taggcatttt 151440
cttccccctc tgacttcttc taggcaccta gaaccaacac tgcctggaca tgtgaaggca 151500
ctcgataaat atttttgaa caaataaatc aacttgcatg gctcctgccc caaactggaa 151560
accccaccta ggaggggtgg gcggggtcat atggtgttca ctcacttacg ctaatgaact 151620
gagaaataac gcacttctgc ccaaattcat gttcattcac actcctctca gcagttttct 151680
gcagtcttcc cagccccacg gaaaattctg cttttgtcag aggaggggat atgcgtgctt 151740
tcccgtgttt gctttaccgc tgggcaatcc atacaaggct actaaactgc agagggtact 151800
ggtgttagca tgccccgtgt ttataaggga cttaaaaaaa tatacaggct tgcatccacc 151860
atacctacca tacttgtgta ctagagatat tctcggggca aaatgaggtg aggtgtggaa 151920
agtgctttaa ggtgactcag agccaccctg ttgcgattgc tgccttcgtg atgactggtg 151980
tggctgcaaa gttcagtggc tgtctttata tcagaataat tctagaataa tttaggagaa 152040
aattctcatt gttaggttcc ttcaagccaa aggaggatgt agtgaaaaga gaataggtgt 152100
tggctgtcta gatgggccct gtttaattag agtcgactgt atcagttgcc aaatgaagcc 152160
aatcttacag ggccatccta tagaacaaat atatattttt tatatttaat atgatatata 152220
tgtgtgtgta cacacacaca cacacacaca cacacacaca tacatatata cagggagaga 152280
tagaatggtt tgcctgctga cttgccatta agtaccgtaa acatcctgga aattgtgaac 152340
agctaattgg aaaacagtct gtccgtgttc atgattcatt gtatgcatcc tctagatctc 152400
aactcaggaa atccacaaag ctgaccaggc cctgctgtca ttttgtggcc agatatggaa 152460
agatataaac cacctccttt cttccctgtc aaaacagttg tgccacgtcc tccccctctt 152520
cctcatcttg actgactccc tcacaggtgg tgtctctgtc tctcctgccc ctgcccccac 152580
acgcaccttt agttacctca gtctttcaaa ttttgctctt tgttcctaag tacagtcttc 152640
cttccaacct ctcgtcatgt catttttggg ccaggaaaga tcctgattat gctataatgc 152700
cactgtacgt gttttaaaaa gaaggaacgc tgtacatttg atattaaatt tggcatttta 152760
aataaagggc tggtaaaaaa atctctgagt gctaatctcc aagaaaggga tggaagactg 152820
gggaaagaga atctacttcc tatttccacc attttaatag cctgacatat ttttttacct 152880
tgcccatatc ttactttcat aacatttttg ttttattttt taaattactc ccatggcggt 152940
agagttgatt tgaactcttg tttttcaatt ttaaatgtac aaaatttcaa ttattttatg 153000
gattaaaata agcaccccag accatcctga gcatctgatc accaatggta agaccattat 153060
ccttctcaag tttcatctac ggtaactggc ttacagataa acttgtggat tacaacctgt 153120
ttgacaactg taaagagcca cattgattaa aatcagaaga ttttcagagt tcagtattta 153180
gactatatgg attatctagt gtctcaatag aaggtaaggt tatggaaatc catttcctag 153240
ttctaaactc tgcaagcaaa caatcatctc cccatagtgt gatatctaaa tagttaatcc 153300
agtatgtcag acaaccccat ttagtaaaca aagactactt gaccatagaa aacatatgat 153360
atatgtataa tatataatcc atatagagta aacatgtatt atattttata tactgtatag 153420
gcacatatca tactatacat atacatatcg cataagagat acagtaaact atatttgtat 153480
tttccaaaat taaatatgtt gcagttcccc taccatagtg aaactgtctc ttctacattc 153540
cttactgcat tccttactat atagtaatac taacactgag cacaatcata tttcaccact 153600
caggatgtag ccagcggata cagtaatggt tcttgtcctc cgcaggagga ccacgggaga 153660
ccagtggctg tgaatgggat gggatttttt tctttcctct aatgaaccaa gccctgggtt 153720
ttattgttgt tgttttaata tacagctatt gagtgttttg tagccacaca cgacaacaca 153780
cacacacaca cacacacaca cacacacaca cacagagtcc ctagcaaggg cagggtgggg 153840
ctagcgggct gggttcccct gggagcccct caccatccgt ttctcccagt gacggcagct 153900
atgtttgaag agcataactg catggtttcc tatgcattca ttcgtgagta gtagctctca 153960
tatattatta aaaagataca ctattattac ttttaaagaa agaaaaggat tgcaattcac 154020
atttacactt tccagcctgt tcttgtgttg tttaaaaaac aaacaaacaa aaaacgatgg 154080
cagaggaaat gtttgcctcc gtagtaggca tcaactttat ttttcaaatc attctgtttt 154140
aacgtgttca tagactgcag ttgtttatag gtatgaggca ctcatcagtg tgaaatagtt 154200
ctttcctttc catatttcct cttatcagaa aaaaaaattc ctgtggtctc ctagcaaaat 154260
acaatccatt ttgctaaatt atttgtgagt ttttataaag tgtgtttaat atcaccaggg 154320
cagaggttca cactagttgc aggattagca agagagacgt agcatgagta gtgtttggtc 154380
cactgcagtg tgttttgtgt gctagcgatc atgagtttat ctgatccttg tttaactact 154440
acacagtgag taagctgtcc tgtattgttc cattcatatt cctctgagtt cattcagaag 154500
cctgacactt cctttgccgg acagattaaa ggggcagcgt gggacctttt gatgatgtga 154560
aacctgcttt cttagtctaa gctccctagg ctatgctgac cactcagagg ttgaactact 154620
atttatttgc cctaaaatga accagaaact tggtcttagt ttccttcctg acacatgttt 154680
taatttccta aaagtgtacg gattttgtag tgggttgttt ttgaatcttt cattttagt 154740
gctgatccag gagagaaagg agatatggaa acatttttt caaaaaatag ctcaaaagaa 154800
aatatgtaaa accatgaaaa acccagaatt gtgctgctgc tttctgtgct aattaaatca 154860
gtgggtgtta ggttgtaatg ataacccttt aactgtgtgg cttatctctc attccatttt 154920
atattatttt cttcctcatg agaaaatcag tgtttattat cacaggtgac aaaacacagg 154980
agaaaaacaa acagtgaggt tacatttaat cactttaagt gggtttcatc tttgcttttt 155040

tgttttcatt cccaagccag aagccgtaaa ccgagcgaga gtgcaaattg cctttctcag 155100
gtgcacgttg ctgagatagg ctgggagaac aggtgtggag cccgtgaaaa gataaacatt 155160
aagtcattct tggggaaacg gtatttagct agacagctga agacggactt ttgaaatacc 155220
attgtgctac tgctgttcaa atattgacta agtgaacctg gaaaggaaga aattttggtc 155280
gcctaacata gaactcgttg tctttttctg tctttaaatg ttatctcaaa gacccaagag 155340
aaggggtagt ttacctaaga aagaaatatg agctttgctt atggagtttc aggtatacct 155400
aatgtaagtt aattaagcaa atacaatgta gcagccttgc atttggccta gcattctttt 155460
atgtttcctg gctgtttctt cgaggagatg acctgcctgt cgggcagatt agaatattta 155520
ctgcagtgca tctttcatgc ctcgctgtga ctctgtaacc acggtggatg tgggaaagcc 155580
attaaccatc agcttgacgg tttacaaaga aataggaagt tcaagttaag cagatattta 155640
ggatatagtt tgccttccac atatttcaac ctgtgttgct gcatactttt taagcttagc 155700
gtaattattc acacagctat gaattttaga agatgtttaa aagcaaacca cagtgacctg 155760
ggaaaggagg gaaacttact ggagcgctta gccaggagct taaaaagaca ttgctagtga 155820
gttttatgtc acatgaaatc tacatttgat aggtcatttt ggtaagtttt tgttgtttta 155880
aatgactcct cttgacacag taaccagtgg tgctgggaac attcattcac attcattcat 155940
ttaagctcat gactcaaata ataacttagt cgtttcctct ctgaaggtag gggaggtaat 156000
gaggagcacc gattaggctc caagatccgt tctgagattc agataaggtg tcctaacaaa 156060
aggtttatgg tgaaatgaaa gagtgagaaa ataattgtgc tttttctagg gtcatgcgtc 156120
aaatgaggct caccaacttt taaaagactt tacatagctt tagataatca cattccctgc 156180
catgtaagca ttgtgatgta atggcatcat catgctactt aacaattaat ttatgcattt 156240
tgtttaaact tcctttagaa tatatatagt ccatataaag aaaattccag ggtcgttttg 156300
gattttgtat aaatagctcc catgtttaca tgtgaaaaaa aattatttat gaaagaaaaa 156360
cagagctttc aatatcctat tttggttacg tctccataaa aactctagga aacagtggga 156420
tcatctgtga aacagtggaa tcaccccaag aacaaactgt cagacagacc gtcctgtcgt 156480
ggcatgactt gaacataacc gtcccacgtg gggacgcatt ccgcaccggt tgctggaact 156540
gacgggggct gcagtgctga atacctctgg gacgcttggg aactgtgccc ctgtttacag 156600
acggcaagcc cttagtggta gggccctgag attctgagaa acataaggtc tgctttattt 156660
aatttcctct cgtttaccaa gagtcacaac ctattttagt aaataaattc aggaaattgg 156720
taaagcactt tactccatcc gttatgcctc ggtcatcagc atggttgtca cggtctctct 156780
ggctcacggg ctgctgcggc tcacagcctt ccctcacttg cctgcaacca gctgagagcc 156840
tccctggtga tgggtgttac tgagcttaaa cgatgtaaac aaacagaacg gcacacaagt 156900
tgtgcaggga agtatatttc ctctaccttg ttaataaaga tttctaactt tagagatttt 156960
ctgtattgac tctggcattc tttccaaata attattttca ccccggggac tacccacaca 157020
ccctgggatg aataaaagaa attatctttc atttgagggt accagcaacc cgctctccag 157080
ctctaatcct cttcatcctc cttctttttt tatttttttt ttttttttt ttttttggt 157140
tgttaaaacc tgagctgctg ccaagctgat cttaatagca tgttcacaaa gacagatgga 157200
ttttttttcct accttcatta gccactgagt gttgttttcc atgatgttct ccagcacttg 157260
cagcctctgc accgagtcat cgtattcgag cggcgcgtcc ctctgcacag cattggacac 157320
gtaggggctg gaggaagagc ggcagttgtc catctctggc aggaggaaag tgtagctgca 157380
ggacccatgc tggacctgat attgcttctt tcctatgctg tccatgctct tccgaaagtt 157440
gttataggct gcggccaaga caagatcaca gctcagagta aagaaaacaa tctgccacat 157500
tctttcttca gtaataaacc agcagcttag caaasttgag ggcaaacaca cgtccagagt 157560
cccgagctgc tgccgtctra aaygcagggc tgctacgctg ccatggctgg gtccgtcart 157620
gaaagtcttc tctttcctct ttttccagta gcaaacctgg tttttactgc tgtgttctct 157680
ccaggcatgc agtaaactgt cagattgcag tgggaagaac agtcctgctc acttgggagg 157740
gctgtgtcag cttttacaga gcagctttca cggtcctttg ttcctctctc cccagatcct 157800
acagtgtcag tatccgaatc aatcactttc ctttccttat atgataagtt gataagagca 157860
gccagacatg tgtagtgggc tgctccgccc tcctggctta gccgttatct tcctgtaggg 157920
ggtcactagc caggcaacag gaaaaatcag agcagaatgc ctgccctcca accaggaccc 157980
atgctgcaga aaccctctgg gaaaaaccga tctgttacag gacccctggg catttcctag 158040
gcaccacccc aattaagtat ttcctagaga gagcagttga tctcttttgt ctgaaactga 158100
tttttgccgt gctaagctgg caaaatatct gaggtaataa ctttaatgtt gaagtacaat 158160
gaaagttcct gtttttttcct ttaggaataa aaatactaca aataggtcag gacttcggtt 158220
tatttttgtt attacaaata aagaggaaga agtttggctc ctgtaaacgt gtgccttttc 158280
agagggaaaa atagattcat tgattttagt tgattcttga accactagcc aagttacaaa 158340
agattttcat ttccgaacag ttggatagaa agatctgtta ttaagtcacg ttagaaacat 158400
cagtttctga gctctgacct ttattcttta aaaaaactcc acttggatat tcactctaaa 158460
aatacactgt actgattaag ttcattacat tacaatagag aaattagaat ttaagtgtct 158520
gtgtagaaag aggaatacaa actttttttt ttttttttt ttttttttg agacggagtc 158580
tcgctctgtc gaccaggcta gagtgcagtg gggcaatgtt ggctcactgc aacctccgcc 158640
ttccgggttc aagtgattct cttgcctcag gctcctgagc agctgggatt acaggcacac 158700
gccaccacac ccggctaatt tttgtatttt tagtagagac agggtttcac catgttaggc 158760
tggtctcaaa ctcatgacct tgtgatcgac tcgcctttgg cctcccaaag tgctgagatt 158820

acaggtgtga gtcaccatgc ctggcccaca aacttcttta ttgtgtcaga atttgttgac 158880
atctcagcat tttgtaacac attatcaatt acattagtcc cccttggtat tagactcggg 158940
caagtcactt ccctgtttta attaagctct aatgttctca tctgtgcaat tcaaggggtg 159000
cactcacaag atttttcacc ttcaatccta tggctctgta agttctacaa gtcacttcct 159060
ttaacaacta aaacttaata cttcagagat taataatatg ttaactcagc agcccaagtg 159120
tacataggga aaaagccccc tgcctttgct gcggtttgtt tatctctcaa ggtacaaggt 159180
ttattattcc cagcgagcgc tgaatagctg gtacactgac ttaacagacc acatctaccc 159240
ataaaagatc tttatttttt actaagctct aaccgaaaga cagcctttcc cttatcaatg 159300
aatagttaac gaacaacagt gtgaatatct gtgactttct catcctcaga aatcagctct 159360
ttttatttgc tgccacaata ctcagaacta catttttatt aaacccagcc ctagatcttg 159420
ctactgaaca ttggaataaa gtagcatgtg tcttcttttg agaaggtgtt tataggcttc 159480
accagacaac caaagggttc tgtcacacag aaaagctgga agacatgctc tggaaggatc 159540
tcattagtag aagaggtagt atgattccac caaggttctg gacatggttt ccactaaggg 159600
aaccaattaa agatgctata cccatccgga cagtgcaccg tcgaagaaag catataggtc 159660
ttaaagatga gacctgtgtt agaaccctgc ttctgtgtga cctccagcaa atgcttccat 159720
tcttggagcc tcagtctccc tagtcataag atggagatca ttttttctct gtagggtttt 159780
taggattaag atataattgt atgtttagaa atatttgttc cttttcttta caggcatgct 159840
ccattaaatg gggatcagtt cttccaccat caaatagtat aactctgcta ttctctgaat 159900
gcaaagcagt ggcagtggca tagggtacaa tttttttatt tcctgtttga aaaagcatat 159960
tgtaaggtat taatatcaca tatgtggttt taccttttt caagattatt ttttgtagag 160020
acggggtctc actcttgttg cccaggctgg tcttgaactc ctggcatcca atgatcctct 160080
tgcttggcct cccaaagtgc tgggatcaca ggtgtgagcc actgtacctg gcctgcatat 160140
atggttttaa aagtcattca gttgtcttcc aggcaaatag agtagtttaa aaggaacaaa 160200
tagaaagagg gcacaccaca gtacttttg tctaccagcc ttgtgtcaga caccatgcta 160260
tgcactgggg atagagatta aggcaactgg gtgtctgacc taaagaagct aatagtgtat 160320
ggagagggag agacccataa acaaattgga tgtggtaaga gagagcagtc ggagcacaaa 160380
gaaagacagt gatgcttccg ggtgcacaat atgccatgtg cagctggcat ggactccatc 160440
cttttcaaat tccctcaagt tccgaacatg gaaggaacag ctctttgtag attcctaaat 160500
ggaaattatc ggaaccagag agtcagggga agctctagta gagccggagt tggcaaactc 160560
tgtccagtgt cagatggtaa ttattttcgg ctctgcaggc tatacggcca ccatcgccac 160620
cactcaaagg taatgcggtg caaagcagct gtgggacagc attaactgag tgagcgtgct 160680
gggctccaat aaagctttat tcgtgatact gaaatttgaa tttcttgtaa ttttcacatc 160740
atgaaagatt ctattttctg caatcattta aaagtttaaa aaccattctt agctcatagg 160800
ctgttagaaa ccgggcctgg actgtagctt gctggcctct gagagcctag gtggtctgtg 160860
ggggtaggag gtgctgggca aggccgtcca cagtgcacgc ggtgtggtga ccgtgtgtgg 160920
ttggcaaggc tgtcagcagt acacacagtg cagcaaccac cgtgtgtggt cagcaaggcc 160980
atccacagtc cacacagtgc aataaccgtg tgtggcgagc aaggctattg acactgtacg 161040
ccgtacagcg accgtgtgtg gtcagcaagg ccatggagag tgcacacagt acagtgacca 161100
cctgtggtcg gcatggccat caacagggca cacagcacgg tgaccatgtg tgatcagcaa 161160
ggccgtggat agtgcccgtg gtgtggtatc catgtgtgat cggcaaggcc atggatagtg 161220
catgtggtgc ggtgaccctg tgtgattggc aaggccatgg atagtgaacg tggtgcggtg 161280
accgtgtgtg atcggcaagg ccatggatag tgcacgtgtt gcggtgacca tgtgtgatcg 161340
gcatgaccat cgacagtgct tatggtgtgg tgaccgtgtg tgatccgcaa ggccatggat 161400
agtgagtgca cacggtgcgg tgaccatgtg tgatcagcaa ggccatggat agtacacgcg 161460
gtgcggtgac catgtgtgat cggcaaggcc atggatagtg cacgcggtgc ggtgaccgtg 161520
tgtgatcggc aaggccatgg atagtacacg cggtgcggtg acgatgtgtg atcggcaagg 161580
ccatggatag tgcacgcggt gcggtgacca tgtgtgatcg gcaaggccat ggatagtgca 161640
cgcggtgtgg tgaccgtgtg tgatcggcaa ggccatagat agtgcacgcg gtgcggtgac 161700
cgtgtgtgat cggcaaggcc atggatagtg cccgtggcgt ggtgtccatg tatgatcagc 161760
aaggccatgg atagtgcacg tggtgcggtg accgtgtgtg atcggtaagg ccatgataga 161820
gcatgcagtg cggtgaccgt gtgtgatcgg taaggccatg atagagcatg cagtgtggtg 161880
accgtgtgtg atccgcaagg tcatggatag tgtacacggt gcgatgacca tgtgtgatcc 161940
gcaaggtcat ggatagtgca cacggtgcgg tgaccatgtt tgatccgcaa ggccatggat 162000
agtgcacacg gtgcagtgac caccgtgtga acaggggagg actggtgcct cggctcagcc 162060
ttctgtgtgg ctgcttacag gggcttacta acgggataga ataggtgctt agagaaagtg 162120
ccacactgaa gtgaattaag gatgccaggt ggggagaggg gccaggaagt ggcctgggat 162180
gcaagtgtgc atggatgggc gctcagctgt ggcccctagg gaagtggaga catggtctgc 162240
caggccactg accaggaagc ctcggggagc caatgggagg cgcttgaagg cattaggtgc 162300
aaagcctgga gttgtgggtg tccagtaaca ccaccatgca gcctgggggt ctgaggccac 162360
ccatctggga ccccttcact ctaaatgagg cttgactagg gggatctcag aagtccacag 162420
aaaatcttgg gtgttcctcc ctgctgtact gacgggacca caagaggcaa gtgagactgt 162480
cagatgagaa acattattac aggttcccaa aatccacctg cctacccacc caatttttgt 162540
ctgtaatagt tctgctgaac agctgtgcat agtgcaattt atttccttaa tactgtttgt 162600

tttctccccc atattctgtg tcggcaactg acatttcaga ggttcccatg tgttctctgt 162660
ggaactgtct caagttctta ttaccctggt tgacgacacc agaaaaacca tagctaccta 162720
ctcccagaaa gaggccagtg ttacaaagaa tctcgtggcc agccctttg gctcagtttg 162780
cccagttgga ggccctaagg cgcaaaccag aaaagccaaa gggcctcctg aggaccgtgg 162840
aagtgggtgg cgcgtggacc catcgctagc tgaatgtgga atgtggaccc atcgctagct 162900
gaatgtggaa tgtggaccca tcgctagctg aatgtggaaa aggacttatg acagtcagac 162960
catcccaggt tcccccagag caatccgtgc agctctcata agcaaccaga aaccaaaaaa 163020
ggatgctaag tcagcacaaa gtggagcagc cccccagcta tgggttgcca aacagaattt 163080
gcttgtgggc cccgtgaccc ctgctgttgt ccagtttaat gctcagcatt tatccagatc 163140
aagggatgga aatggggcca ccagcctgac ccaggcccgg ggtcgttttg cttttccaac 163200
ctgtaccatc ccagcaatgc attgccagcg tgcaatttga aaaagccctg ccgagctgaa 163260
aaacacatgg gaagggctca gacacactta aaggcacatt gctgccctgc atttatacgg 163320
cattttgtgc tgacatcgtt ttccatcagg cctgggcagc ccctcctgag actgtctccc 163380
gcctgccgtc ctcagcacgg cctgcccggc tacagtctgc tttcctccca ctgcccctgc 163440
ctgcaggcct tggaggcggt gactgctgca gacttatttg ggcagcctgg ccttaatttt 163500
tggaaagtgc cttgttgatg tatgaggaac ttccacggct gaaacagtct aaaaaaatga 163560
agctgggaca ctatgttttg attttagcca tttgcagaca gaggggcaca ctcgggactc 163620
ttgggcgcct ggcacactaa gctgggaggg acttttgaga catcttggcc atctaaatca 163680
gtcaacatgt ttatatatac aatttaatgt tcagtataca gggaaaacca ttagaaggtt 163740
agctgcacat aaaactgttg ttaaagttat ttttattact tccccccaca aatcgtatgc 163800
aataattaat aagaactaga gaaatagcca caactggcac aacacctgcc cctctgccaa 163860
aagaaaaaaa tcttcttttct gaaggcaggc tccctatata gtgattcctt tatatgcctc 163920
ctggaagatc tgtttcgact ccattttgat atatgttgaa ccagatttga agacccacaa 163980
atgcagtcta gagccatttt gcaaaagtgt tgctgcatca accatttcca ttccccagtg 164040
ctgctcatca tgttacacta gtgttaaatc ctgactttgg aatgcgagga aggacagttc 164100
cagccatggg atttcaaaaa agtaccaaag gaaagcccct tcaagttacc gttaagacag 164160
aagaaaagga agaaaaatat aaacacacac gtataaacat gtaaggtagc tttggtccct 164220
ataacagaca aggaaatcaa ggctccgtga agagagagac aagaattccc ttagccaagt 164280
gcctgtgtgt gtctgtcttt tatgttaatg gttatgaatt taaggagaat tgaaagcaat 164340
aattttgccc ctctttaaca tggcaaatac agcctgcttt agagatgatc agcaatcacc 164400
atttagtact ggccgtcacc tctgtgcagc acaaacacac atcccgagtg acagaagcca 164460
tttcactgcc agagactctt agcggccttc agttctcttg agctggagcc actgggtctt 164520
gtatgaaagc tcaccagaca tctcatgtgg acctcgggca tctgagccgg gaccatccta 164580
ttacaagtgc ggaaaccaga tcattaatgc agagctgaat tcaaattgtt acttgctagc 164640
ttaggaaaga atccttggaa atccaacata ttgtctaaat ggatcagtta atcttactat 164700
gtgcattcta catacccttt cattgtttgg gcttaaataa cttttctgct ttgtctggtt 164760
taatttcatc caatgtggat cgctggaaga atatgatgta tgttttagaa tagaaacagt 164820
tctgagatga agttgagcac aatttcctgt tctagttgca attaaatata aatatagcat 164880
ttgacataaa atagctggcc cgatatattt agagtacaag ttaagtgtca tccccttaga 164940
attgggcatt gactccgtag aattcccctt tgtacaaggt gagcaaatgt atattttgtt 165000
aaaaataagt atctgactgc caaaacggac agaaagctct ttgccatatg tgttttcagg 165060
ccatttcctt tcctgggaaa cagccatttc ccccgcatta tagttgtgtt ttcatttgcg 165120
ggtagataga gtaagcgcag gagttaaagg acgcgggcct ccacagccaa ggccttatct 165180
gggacaatta tctttctcct tgcagctgtg taacttctgt ttgacacaga accacagaaa 165240
ccctgttagt gggaaggatc acagttaata ggagaaaaat cttcattgtt catgagactt 165300
ctcaggtgct tggcattctt atttaggtgg cttaaaaaag ttccaagtac tcattcattc 165360
taacttatct gtgttcattg tgaaatcgtg tgtgaatgac atttggagca gatggattgt 165420
tgtttttttt ttttttttttt tttaacaaac ttaagagatt cccgaatctt tcacagtttg 165480
tactaccgca aaccagcata acatctgcta aagaatttca tattttaaag ctgcactgta 165540
catcatatgg aaccttaagg actttgaagg gaagagcttt ttatttactg gtagcttggg 165600
aaatatccaa gtaactattt tttaagaaaa aaaaattcct tgagttttta gaaatagttt 165660
atataactgt tatgctgttt gatttttaaa tattttcatt ctctagtatt attatggaat 165720
attttatctt cccatcaaaa aaatgccaga aggtcaagat agaagtcaca acattaaaag 165780
ggagtggata caattgtaaa acaatagatg agtacatttg cctgataata tttttgccag 165840
taattctgtg tcctgttttc tccctgtaga atgaaatgct aaacattttt ttcaatggat 165900
tgatgtcagt gtttactaac atgacctgtg ttaagtcaaa taaagtattt cctttgacaa 165960
acaccatatt tcattagtgg ctttgaggtg ggcttatttg ttataagtca cattaaatgt 166020
tcccaaatcc atttcataaa tgttgtcgag atctcaaact ccgttgcttc taaaaaaata 166080
tgtccagtct ctttgtcata accatcctaa taaagatcta aatttcttag agtgaatttt 166140
catttgaaag tggcttaatg ccagctagat taattcttgt ttaatctaaa tttataaaat 166200
ttttatctta attattgaga aacctttta aaaagagata aaaatgtcat atgtgctatt 166260
tacattaaga tatattatct ctcttggtta taggttaaga taaataaaat tgcttatgtc 166320
aaagaagtaa aaaaaagtcc atgacctcct tttggtatcc ccatccatct ggcggactta 166380

atatgaaaaa atcttcctgt gggaaattag gcttgattat agagttacaa gtacaaaaag 166440
tagtttttga agaattataa taaatagtta cacataaaag gaagtgatgt ttgcttgaag 166500
tatataaaaa tattccttgt cactcttgtc ccctcatgaa tcttagttgt ctgatgatgg 166560
ttcaagtctt tcctaataat ccagaatgta tccctccact ttttctctta aaaacgctat 166620
ttcaagcatt ttctttggta ccccattaat aataaagcat acttccccaa aatgttccat 166680
ttcaagtaag gggtctaaaa gtcaaagacc gactgataca aaagagaaaa gtaaattgta 166740
caaagactga agagaggatg cagtattaaa cgtaccaagt tcttgacatc ggtttccctc 166800
aagaaaaaaa aaaatgagta acgtttttg aaagcctgaa actattctag taaaatattt 166860
acggaaaaaa taatatgcgc tctcctccca aatcctgatg cgcatttaaa tcacctttt 166920
tatttataga tcaaaaatct tgcttgacta caataaaaat taaaaaatgg tacctattta 166980
agaatgcaag tatcaaatcc acttgtaata ctcactagct ccctctgctg atctcctatc 167040
aagcgacagg caaatctatc catgattgtt attacaattg ttaatggaaa tgataggtaa 167100
tttaggacct acatcaattg caactaaaat acaagctaca atgctttcat tttaatttta 167160
atgcaaaagc acatcacacc atatacagat gttaaagacc gacgtgcaca cacacagtga 167220
aaaaatattt ttaggcattc atttagcata catagaccta ggagctgtct ctgtatcctc 167280
aggtgataag gttactacta ttacaacagc agaaaaagag gtctgtactg tctgtctcca 167340
taaggagcca atttagagac ccaatcctgt tcaccccaag cttacagtct aacgaggtga 167400
acagatgtcc catctggatg cacaagcact gctggctaag gccctgggta gtgcaggagg 167460
gagcccccac acgggaagcc tcccaaacca cgtaagggct acgtgaacag caagaatagt 167520
ttcactgttt atttagatcc acactgttac ttatttaaga agaacatact ctgccctttc 167580
tccctccctg aagaaagacc aaaactgagg gaaattatat tccaggctga gaaaattgcc 167640
tgtgcactta aaaaataaat aaataaaagg cgagaccacg gaagttaaaa taaattaaca 167700
ataattgagc caagagggag gagatgggtg agtcggagat gcggtctgga actagctgct 167760
gaagagtctg cttaggaatt ggggttgtac cctggacata aagcatttgg ggcgggggag 167820
tgtcctgatg tgactgagaa aggactgtgg agtgctgtgt gcagtagggc ttagaggagg 167880
tgtgagtaga ggcagagaga ccagagcaga agctgctaca gtaattcagg ttagatatta 167940
tagtggcctg tgctagaata ttaacatcag gctcatgatg ttaaagaggg gtgatcaata 168000
ggccttctag atggatggaa tacagggagt ccaggaatgg gattggtctg gtactgggac 168060
tgactcttac acttaaaaat gctaaaataa aataacttgg caatagtttt taaagcatct 168120
gtaaaatgac aagataataa acacatattc tgctcaaact tatgtgaagg caggaatcct 168180
gtgaactatt aaagagcttt gccatcaagt catttgccaa acctggccaa cttaagccta 168240
cttcaaggcc tgagtggttc agacagaaga caaaggccag gacctaaaga aatgggagca 168300
tctgatgaga tacctccttc caggaaggct ctaccccagt gtcagggaag cagaagtaaa 168360
cctgcccacc ccactctcca gagcagacaa gaaaacatgc ctggatgtta aacagaacta 168420
aaagagggga gcccatccct gagaatttaa ctacaagctc acccttttgg gttttacagt 168480
acacataagg tggccagaaa aaaccacaat gaattgttct aaggtggtcc caggctgatc 168540
atcttattcc cctaggtttg tggaagaagc aaatgaaaat cctttctggg agaatgcact 168600
ttcatcatgg gtctcaaaac attcttacaa tttcccaaga ataatgggca actcacagac 168660
aaaaataaac acacaagaaa acatagtgct attagcaaaa atcagcagaa agaagaaata 168720
gtaaaaacag accagccaaa aacatctatc cctgctgtat tggtttgcta aggctgcggt 168780
acaatgtgcc acaaaccggg tgcctaaaac aatgggaatt tattctcaca gctctggagg 168840
ctagaagtct gtaattgagg cgtcacaggg ccatgctccc tctgaaacct gtagggggtc 168900
cttccctgtc ccatcctagc ttctggtgtt gctggcctca tcgtctcatg gtattctccc 168960
tgtctacacg gccgtctttt tataaagatg cagtcacatt ggattaagag cccaccccac 169020
tccaggagga cctcctctca actagttaaa cctgcagtga cgctacttcc aaataagccc 169080
acatgctgag ttgctgtggc ttaggactta catctttcta tcaggaatgt aattctatcc 169140
ataacactta cgatgtttaa agatgaaagt aaaattttga aaacacctaa aggggacaga 169200
aaactaaaga agaaaatctt gcagatttaa aatgaaaact tatagacatg aaaaatacga 169260
tggttggaat ttataattca gagtcatgtt taacaagatt agacacacct gaagctgaaa 169320
gataagtgaa aacgagtcat cccaggtgta ggacagagag acgagataag gaccaggaga 169380
gaaagtgaag gaagacctgc aggatggagg aagagtgtct gacaaagtcc atccaaattc 169440
cagaagaagg gagagagaac agggcaggca atatccaggg agatagtcac tgaagctttt 169500
cccaaagtga tgaaagatat caagttacag attcgaaaaa cggcaaaaaa tgacaaacag 169560
gataaataaa atgatgagat aaagcagtaa gagggaagtc aagaattact ggttctgcag 169620
tttctggcct ggcagctgca cagataggat tctcaagggc tcggaagggg agaggtagca 169680
gagaggcaat gcatgcagct tgcacacatt cagtttaaat tggctatgag acttccggtt 169740
agagttttcc tattgcatat ttgggtctga gctcaagaga gccacctggg ttggaagcag 169800
caggaagacc tattttcctg attaatctca atgccagcct cattacacaa tcttaactaa 169860
tattaaacag tatatgaaac aggtgaagaa gaacagctgt ataaattgca taaagcttag 169920
caatgtgggt ttttctagac aaagttaagc agcaaagcag ctccattatg agggaccctt 169980
ggccacggtt tcacaggtgc aggttctgca gatcatggca tgttgtcctg ttctctggat 170040
tatggctcta gaagagataa tgataaagaa gacccagggt ggtcagtaaa aaggtcctac 170100
gtggtgtcta tacaatgttg caagtgacta aaaatgagta aaacttacaa gatataatta 170160

gactaattat caaagaaaga ttgctttagt gaatgagaca gtttagatcc attcccttgg 174000
gaaattatgg tggtcagccc tccgctcggt ctcactttta gataccagaa actatatgtc 174060
cttgtgttgg cagagctgga ttgtctgtcg ccctctggtg caatcctgca ttagtaaggg 174120
aagtgttttt ctggggcgtt ctaatgaaaa gtgcttaagc atttgttttg gtgcccagat 174180
aatgtgactg tagttagtat gtagtgtttg gacttttgc tcatgctttt gttgttgttg 174240
ttgtcattgc agaaataaaa ttaacccctt aatcttatgc ttaatgtaca caccaagtgg 174300
tttgcatatt atactgagaa aataaaaaga ttgttttaga aaaaccaaag gacaccaaca 174360
gctctttaca gccccaaagc aggtgtcgcc agaggtcaca ggaggggttc ttagttatca 174420
gcaagggaaa ctgaggcttt ctcgtttatg cagaagtgga atttattgaa taatattaag 174480
ggggctatgt cgccaatgcc acagtcacac tgcccacaca gaactggcct ggcgaggtgt 174540
tactttgacc accattgctg ggccaggacg ctgccaccaa ggccgtgccc ctgccagaaa 174600
ctaaatgtgg ctgccccatc cctggccctt tctgtcagta gggtcaggtt caaactcctg 174660
ggtagtcagc ccagctctca ttgactcagt ctgaacagct gcctgttccc tagaatccac 174720
atgcgctggg acaatgggaa gtatcggtag acgctatggt gggaagatga ctctgtgtcc 174780
accaaggttc ttgggctggg gaatggtctg agcatatgac ggcctcagac cccagccaac 174840
caaagggaaa ggtctcccct gtactcacga agcctccacg atgtccatca gcactttctt 174900
cctccgttgc agtgtaggtc agcccttcgc agatgctcac aattccctga tacagccggt 174960
tgccctttgt tgtgttaaac tgaaagaatt tcagagttgg ggccaggcat ggtggttcat 175020
gcctgtaatc ccagcacttt gggaggccga ggcgggcaga tcacgaggcc aggagttcaa 175080
gaccagcctg gccaacatag tgaaaccccg tctctactaa aaatacaaaa attagctggg 175140
catagtggcg tgttcctgta attccagcta ctcgggaggc tgaggcagaa ttgcttaaac 175200
cgggaggcag acgttgcagt gagctgtgat catgccactg cactccagcc tgggctacag 175260
agcaagactc tatctcaaac acaaaaacaa aaacaaaaca aaacaaaaaa aaactcagag 175320
ttggagaagg actcggacaa atgtcatatt atagaggagg aaaaagatcc aggaggcaga 175380
aagacttccc tgagggccat gatggtagtt agtgcatcca ttaaatacaa gtcttctgct 175440
tcttattcct gtaaataagt ttgcatttaa catttttgta cattaaacgt tactgattca 175500
tagtcaatga ttatggtcag ccctccacat ccgcaggttc tgcatctgta ggttcaacca 175560
atcgtggacc aaatatattc aagaaaatga aataaaaata caacaataaa aaagtacaaa 175620
aaatcgagta caacaactat ttacatagca tttacattgt attaactatc ataagtaatc 175680
tagggatgat ttaaactatg tgggaagatg tgcataggtt atatgcaaat actccatttt 175740
atataaagac ttgagcatcc atggattttg atatccaagg tgggggtctt ggaaccccac 175800
aaataccaag ggacaactgt gtattatttt cataacccat ttctgcctag tgttccatta 175860
gtggaatgct aaccatgtgg gaattattta tatcctactg ttcaaggtca tcaccaaggt 175920
ctgatttttc acacacacac agaattgcaa cctccagcat aaatggggat gaatttacta 175980
ctaacatgta gtttccatcc acaaatccaa tgtccctatg ctatttgtaa ctgtggagcc 176040
aagagaagct gttgaatcat gtggtgaata tgatcaagaa ctcaagatta gggataaaag 176100
caatcattct gttattcctt tttaaaaatt attagcctgt aatttaaaca tcaggatctc 176160
atgtaataca gaacaatatc ttctgacatt tttacaatac tagtattctt acaaaacaca 176220
gttaggaagt tacatgaaga aaacacccag actgtgtgtg gctaaatctt tagtacctca 176280
tttccatagt cttagagaaa gtttaaatta tattgaaact tttctcaact gctatcttaa 176340
tgtgttcagg ctgctgtaac aacatatcat tcaaactggg tgtcttataa acgatagaaa 176400
tttatttctc acagttctgg aggctgagaa gtccaatatc caggcagatt ccatgtctgg 176460
tgagggcctg tttcctggtt catagatggc gccttctctg cgtcctcaca tggcagaagg 176520
ggtgagggag ctctctgggg tccctttttat aaggacacta atcccatttg tgaggatttt 176580
cactctcatg acctgctcac tttctaaagg caccacctcc cagtactctt gcattgggga 176640
ttaggcttca acatgaattt gagggaggcg caaacattca gaccatagcc actggtcaac 176700
attaggtaac ctgcagtgct tggctgtggg atgggaagcc tgtgttgtaa aggacgtctg 176760
agtgggaaca ggggtctcaa gctgccttca catctaacgt cagcacacta gagatggaca 176820
ttgcagctgc aacctactgt gcctgtaaag catttagaat tacgccttgc atacacaaag 176880
tgctcaataa atgttaactg ttattatggt tgggcatcag ccactttaat tatctctttc 176940
aatcctcata gtaactcttc aacataggta gccttatttt gcagttgagg aaactggagc 177000
ttagcaaagt ttagtgacgt tgcagagcta gagttcaaac ccaagtctga ctccaaagtg 177060
catctatctg tgtatttgct tatttaacct cagacacaca gaatcggatt aattagagtc 177120
cttgattcag cacacgttct cttcattgat ccttactcct ttattttatt ttttaatgct 177180
atttttgtt tgtttgtatt taatagtaag ataaacactg tgaactcacc acttacctct 177240
catcatgaga gcgctggtgc ccacctccac ctccgagttc cacatatccc attaccctgc 177300
cttccccgtc caaggaaacc actgtctgga atccttcgtc attcaagcct tttcacagta 177360
tggctctttc cagcctttta tttctctact gtttcgcttg gaaactctac atttctaaga 177420
cagtgtggtg cctctgagct ctgtggcttt tgctcctgct agcctttctt cataaagtct 177480
ttcagcccca caagtgtcgc agcttttcaa agcctttccc atcctttaag gtcctacttt 177540
tcttttccat gaagtcttct ctgggccacg atgactgggg aatcctcact gtcttctgaa 177600
gttctgcacg tacttactct gcacatagtc ggcggtgagg tattcatcac attgaaatcg 177660
agttacatgt ggtcctgttc tatagtcaac caaaactcct ggggtaaaaa tgctgctttt 177720

catcttggca atctctatcc taaccagcac agtgcctcgc tgaatattag aggcctgaga 177780
attttctttg ttttttttc agagtgattt ttttttctct gctttatttg atactttgaa 177840
gcagcacaca tttcagtttg ctttatgctt gatttttttt tatttcttct aaacaaacga 177900
gatacatgtg cagaacgtgc aggattgaca caaataatag ctggcagagt gtcctaggaa 177960
agactcctca gatgttataa ataatacaca aacaaaaaca cacacaaata tttactgaag 178020
acttttcttg ctctgcaagg cactggctgt gtgatgcaga ataaaaccga caaaattctt 178080
gccatctggg atgtgcatgt tatgtcagca cagggaagag atcaagtgtg tgtgcatagg 178140
acatcaagaa tacaataaaa caaagtggac aaaaggaagc gagggtggtg aacacaggac 178200
acctgaatga aggacagagt tgttggaaaa ggacccctga gtgcccaagg gaggagctgg 178260
cctggagtag tgagggcaag gtgattgcaa atgaggcctt ggtgattgga aatgagctca 178320
tccccacatc ttataaatag ttctccaagt tatccgaggc aggttattct gtggcaaaga 178380
cgcctcagct aactggatgc agaagagaca actgaataga gcctcatggt ctcggagtct 178440
ttttttttt ttttttaaga catatctttg gcattttgta cctaccttct gttctaaatt 178500
ttgcattttt actactttca agtgggtgga ctttgttgtg gtgggtagtt caagattcat 178560
catacaaatg tgattgtgct tcgaaactcc caccagtctg acgcacgcat gggttttctg 178620
gcaacatttg ccatctacag cactctcttt gatcaccttc atcatcttcc aacattcctg 178680
ccacagtcac ttcccagaaa cttgctaatc tgtaatagaa accctcagat tcctatggtg 178740
aatttgtaat caaaagtcac atattgattt caaaatcaat acacacttta aaaataacac 178800
tacagattta gcagctcagg gaggaaggaa accgtaagtt catctggtgc agctacccgt 178860
ctgggatgtg aattcctcct cttcatgaaa tgtttacatt catatcacag tctagggttt 178920
agtgaaccat aaaaagctga aagttaatgc aaacagaagt cgcccccaaa acatatacca 178980
actgatttaa aaggagacac agcagatgga gattattgtg aaaagaactc ttactggaca 179040
atttttttgt tattttaatc tctgcttatc ccaattcttt tagctgcata tactgagaca 179100
cttcacatct ataataaact tggtaccaga acacaattca ttccagacct aactctttta 179160
gatcattata accgggggag gaaaaaagtt aaaaaggctt atctatctta agaagtattt 179220
ctcagtgttc gctacacgtc acttaatctt ttccaaaatt tgacaatata caaagcagtt 179280
tgtagtgact tttcatagtg actctacaat aaaatgggcc tgtcctcctt gcttttccaa 179340
atgcagtcat catctgacaa ggtttagcta tttggggaag tccttgcttg caaacgtagt 179400
tcttttgcca aacaggtttg gtcaaactgt gtcccctagt tgcacagtta ccccatattt 179460
gattaacaaa tagcaaaaca gagataatct cagaaatatt caagagtctc aaaccccaaa 179520
taaaatatag gcatcctcct gttgagtcga attggcaatt ttgattagca aggctcatga 179580
agcagtagat atcccctctg atccccatcc cagtgcgagg gcacagtgag ttgtattttc 179640
taagtataaa ctattctcta gcagttcggc tggagtattg ggagcaaaac tgtatttttc 179700
taatattttc agactaagac agtgtctctg ttttctggac ttttccgtgg caaatgaagg 179760
atttatcagc aatacaaaga aagttctccc agtgggtact ccacggggag aggagctggg 179820
gtctcactag tgcacagcca taaaagacac cacaagcata ttacacgtga agcaggatcc 179880
gtgcccacca cagcagttgt cccaggagtt tcctgtttga atgagacact ttgggtggat 179940
actgcaggga gggagaagct gtgtgtggcc accacagctg gaagcgtggc ctggtgccct 180000
cacagctgtc tgggagcccc ttcccgggaa cgccggcttt tcccgggtgc accattgcag 180060
ctggagccgt tgtcggccgc ctcgaaaaca tgcagttggg ctgctctggc aggcttctcc 180120
agccctcctc ccaaggttta cctctctaaa tgtcaaaagg gagagaatac tgtatttgtt 180180
tttcctcta ctgaaattta tttgtgacat caggcatcac tttcacctta gtcattttgg 180240
ctggattccc atactcaatt aaatatcctt ccttccatat ggcccatagg aagagagaga 180300
aattacatgt aactggtctt tcctcctctt tataaagtct ggtggctgag caacttggcc 180360
tgtacttcct tcatgaccca ccatcccatg actgcagggc agttttaaac acagcagctt 180420
ggtttctatt gcacggaagc tggccaacag tcacagtgtg cattttttcta ttgcacctcc 180480
ttgtgttaac ccaagttcac tcacagctgt aactacagaa gtttttctga aagcaagtga 180540
agccatcctt cttttattga gtttttgagc tagggtctca ctctgtcacc caggctggag 180600
tgcaatgatg tgaacatggc tyactgcagc cttgacctcc tgggttcaag tgatccttgt 180660
acttcaacct cctgagtagc taggactgca agcatgtgcc accatgccca cgctttctga 180720
ttttttttg tagagacaga gtctctctat gttgcccagg ctggtcttga accactgggc 180780
tcaagtgatc ctcctgcctc agtctcccrg agtcctggga ttacaggtgt gagccaccat 180840
gcccagctca tccttccttt aaaaccggca gctgggcaat aatacagatg ggaccaacta 180900
agtttctcag accactcagg gaagctagtc ttgcatagac aaaatataca ccctcttacc 180960
tgccccacct ttaaggctgg tccccagggt ccgcgctctg tcctccagcc tccacgcttc 181020
cctgtgacta gcctctgtgg tcaaaggtgc ttgctgatgc agcctctgta cagcctccat 181080
gcagtgcgtg tctttatgtg gaggagaccg cccttctttc agcagttatt gagcatctac 181140
ccactctgtg ccggtcatag ggcttagaac tgcatgtctg gggggaattc tgcaaagaga 181200
gcctgaaata aaggcaaaca gtgagagacg gccaggagaa accatgagca ctgcagtgag 181260
tatcaaggga caaagctgaa aaaggaagac tgaacgctga gcttcaagcc attcatttct 181320
atgggccgcg ggagcccttg aaagtctgtg ggcaagtttt ggtgagatta agctggtagt 181380
tctgttcagg acaggttgaa gggatgagag attaggacac ttaccacctg aatcctgtcg 181440
ctggctttag tttaaaccac ccgtaatgta gacatcctga cttagaattc cctgtgctgc 181500

ttcctttctg atggaaacag ctctgctaac agagtgcagg ctgtgggagc cgagccccgt 181560
tgcaggcagc ctgcaggccg cagtttcctc ggcttaccac ccagcgcttt tcattcggct 181620
cagcgctagg gacctctgct tccacttctc ggtgttggaa attgccattt atttttgctg 181680
tcgatgatct gtattgactt ggcctgagta tgcgtgcacg tctctggtgg tctgaattat 181740
atagaccaga agggtgtctg atgccgcttt tataaaaaat aataataatt tgaaaggaaa 181800
aatgactcac tgaagtctgg caaatacaga gccctctctc tgaatcgact tctcacttgg 181860
ccatgttgaa ttccaactgg gtgtcctcag acatttctat cccaagatct actcctggct 181920
tagaatctgt tttgttttgt cttatttcag ctcatggttc ttgttccccc agctttatgg 181980
ggtataattt ccatacaata gaattcaaca ctttcaatgt gtggttggat ggcttttggc 182040
aattgtatac agttttgcga tcacccctac actcaagata tagaacactg tttctcgtct 182100
ggtgattgct ggacattgaa ttctttccag ttttcactgt tatgaatctg actatgattt 182160
ttggcatgga tttgtatgta gctataaatc acttggtaat ttttcagaag aatagcagtc 182220
ttggggcctg gatggcttat tgtggtctca aaaagttcct gatgataagg ttgcagcctc 182280
atgcttcttt ataagaatgc agtattactt gcaagggagc ttgggtagat aagaaagcaa 182340
gaaagtccat gtggagaccc tgtccagaga gcacagacat ggactaagtt aaaggatggt 182400
aaattagcaa tgcccaaaag cacatggagg agatacttcc cctcctgact ctattggtga 182460
tgcagtttat ttgtctgagc tatctgagca agtttcctct cacttacgtg ctggggacag 182520
cagattccaa tgcagagtcc ttagagctca ggctcccctc aacctgacgc atctctcaac 182580
catttgtctt aagctgtctg aagtcagctt cccatcttgg ggaggtagaa gtgaaagggt 182640
ttccactttg ccaagtgagc gtatatgggg agactgaggg tgtggagttg atgatggttg 182700
tggggtggct gacagtgtcc acagggctag tcttgaggca ggctgacact ggggccagat 182760
gggaccactg tgcctcctgt cccctccacc ttctcctagt ccaggaaggg aatagcagca 182820
gctgctctca gtggggcatt ctttttccag agacaggcca gcccagcagt gatcccttga 182880
taaagcaagt caccgttatc agagcaagaa ctatacattc acttaaaact ttttttttt 182940
tttaagtgta aaatgggact gcaacaaaaa gaaaattgtg cttaggagaa tgtccctcag 183000
aaaatgtact ttatgattgc gaggaatatt tgccaaggtc tttggggtag gctgagcccc 183060
ttcacctccc tggggacatg ctaggatggc aagagaggat cagacatctc ccagggaggc 183120
tgtgtccagc cgggctcctg gagtggcgta agtctggttg aaccagcact gaactgcctg 183180
agtccatgtg aacgcattga actgttaaac cgtgtctctg gcggccacat ctccgggctt 183240
cacccgctgc tctcccctgt cctgcaggta caaagtcaat agtcaacctc agttttgaat 183300
gttacaaaat tattagcctc tccatagttc ttcccatggc ttctcaccca agccttctgc 183360
tcctctctcc tctctgccca ggtctcacca gctgcccttg ggccaggtca ctgcagtgtc 183420
tgccagcacc acgacaggca ggctggaggc ccagttctca cagaaaagac tcgaaagggg 183480
gctttccatc ctttatagtc tacctgctac ttataggcca ccaggacaaa ggatcaaggt 183540
ggcaaggcag aaattgcagc acagagcgaa tggaaaggca gtcactgaag ggattctttt 183600
gcttttacaa gtagattttt cttaaacaat cactgtatga aaacaaaagt acaaaattat 183660
taaaacacct ggatgatgaa ttgacaacaa gagtttttct ggaacatcct cctgtgggct 183720
cggggaagac agttttttc tgtggtgata gatggtcagg aaatgtagtg acatagaagt 183780
gaaggcattt tacagagctc accttaatca atggcttttt cacttattaa gttttctttt 183840
atttttcct tcttcaaaaa cgactgatac cttaatttat gggaattgtt tccagtaaaa 183900
attgggacaa tgatagtgag tggagaatat ttatatgcta tacttcctgt cttccttcat 183960
cttttattac tgaggatatt gacatgaaaa caggatcttt gtatccaatg agttcatcga 184020
cggccgattt cccaccagaa attccaggct ttctgacatc agcgtgcatt gctctgcatg 184080
tcacttggag caccggcatc tggaaatgat gaaatcctga acaacaaagt ttgttttcag 184140
gaagacaagg cagtggggaa gggaagggtg ctaagcttca gtgactgcct actgtgtgcc 184200
aggcattttc atcttccatc tcgatagaat gtctaactgt gctctgagat gagaactata 184260
aatagctggt cagccaaaag ttttctgctt tttcttagtg atctcaagtg tttccatgac 184320
acgtgctgca accaaataca ttatgtgtaa attgccaaag acctgttgat ttccaaacca 184380
ttatatagtc atgggaatgc ttgtatacct gaattgtcat aaaattgatg agatgcgaag 184440
atacagcaga atatatcaga taattctgca gaactcttat tatggaaatg aaaataattc 184500
aatagagaag tctcgattca taaaagacta gttttactct aaagtatcta aaagacatgc 184560
attaaaaaga catggcactg tccccgaaat gatcttgctg tgttgcattt caaatggtac 184620
cttcattttg aaactttgca cattagcacg ttctttataa tagcaaaaag tgggggagtg 184680
aggaacattt ggtgccggaa gaatgattag gtaaagcaca ccaagctgaa aaaagtattt 184740
ttgcagagcg ttttcaagag catggaagag tgttataatg ttaagtgaac aaaaaaaaaa 184800
aaaaatacag atccaactat gtaatcatta cacatagaaa taaaaatgag caataaagcc 184860
aggatgtcag tgaggatgga gtggagggaa tgtcctaaat gtgcgttggc ccatcatcac 184920
ctcatgcatg aagtgaatgg aaacatttgg tttatgtttt ctggaatgtc tcataagcca 184980
ttgtaaccaa aaactacacc atgaacaaaa agcaaagcag gccctgcagg ccctgggtgg 185040
gaagctgagg aggttggcag tttctcaaac tcatgtcaga tgcccctcgg ccactagaca 185100
gaatctgctg ctatttgggt tctggttgac cagaggccta atctggaatc tggttctaaa 185160
aaccaatttt tgttataggg cttgctggat acaaatctgc aatgagacat tgtcacaagc 185220
aatagcttaa gaaaaacata aaggaaaaaa taataataag tttttggaaa taagcctgga 185280

aaagcagttt attgccatct gctaactcat ttgattcttg cagtaaccct agggtaggta 185340
tgatggtgat ctctgcttca aagatgagaa aatcgaggct gcctcaggtc acttgacctc 185400
ctcacaggcc agtggagact ggcttcagac tcgggccttt ggacctcaag gccctggtct 185460
tcttttgttg tttgtttgtt tttgtttttg tttgtttgtt ttcctgagat ggaattttgc 185520
tcttgttgcc caggctggag tccaatggcg caatctcggc tcactgcaac ctccgcttcc 185580
tgggctcaag taattctgcc tcagcctccc gagtagctgg gattgcaagc atgtgccacc 185640
acacccggtt aattgtgtat ttttagtaga gacgggtttc tccatgttgg tcaggctgtt 185700
ctcaaactcc tgacctcagg tgatcagccc gccttggcct cccaacgtgc tgggattaca 185760
ggcgtgagcc accacgagcg gccaaggcct tggtcttcct atcgcatttt gacaccttgc 185820
tcagtacgat gagtagttaa aatcactgtc attggctaca tgcctacttt ttatagtcac 185880
tctacttatt gtggttttgg ctacatccta gttgaactct agggctagtg tttattaagg 185940
tcttgatctc atatggcatt tgtagacaat cgaatgttga gtgataagcc ctggtaacgt 186000
gatttctcac tgctggcccg tgaagccatg gaaatgttcc catggaaatc accccatgtg 186060
tggaatgaat ggtcagtgga acataggcat ctttctctcc tgtcctctag gttttaaaat 186120
acctgaatgt cctgaaatgc aagagtatcc taagagcact ttagaaatat ctttgcggtt 186180
tctttctggt gtgctttgtg ggttgggtga ggtaccgtat tccaggacac gtggcccttа 186240
gagaaacaaa taatttcctt tcctcctgct tcagtgttat tggtaaagtg ggaaggtagc 186300
cccaagacac tcagctcctg cactgcattt ggatagaagg gcgttcaaat tccaccaggg 186360
acaacttcgt ctaaccccct agaattcctc attttgaccc ttggcatact ctatatttgt 186420
tgaaatacaa aaaaaggagt tgaaagtgag tctatctata tgtagtaggt atatcgtgtt 186480
cactgtaaaa ttccttactg tatgtttaaa atttttcaga atacaatgct gggggaaacc 186540
tatggaacag aagtagggaa aaaattcgac aacgcaaagt gagagtggga aaccatgtga 186600
agctctgtta gagtatcatc actaatctct tttttcctta tacctatatt catgaaagca 186660
aatagagaac aatacaatat agtgtaacac cgtgtaccca tcactcagca ttgctcaatc 186720
ttagttatca ttatggttat tattattatt attatttgag acaggacctt gttctgtcac 186780
ccaaactgga gtgcaatggg gtgatcctgg ctcactgcag ctcaacctct cgggctcaag 186840
tgatcctccc acctcagcct cccaagtagc tgggactaca cgtgcgtgcc accacacccg 186900
gctaattatt tttggtagag acagggtttt gccatgttgc tcagggaggt ctcaaactcc 186960
tggactcaag caatcctccc accttggcca attttaatat tttattatag ttgtttccat 187020
tttttgtgtt tttcataaat taaatcttgt aactattata tatttcacag aatattataa 187080
agttaaagct ccctttgcat ctttccctct ccaattccat tcttcctctc tctctaaaag 187140
taactgctgt cctgaattta acgatgattt ttaaagtcat ctaggctctc gtttttcttt 187200
ctttttttt ttttttttt ttttttttt tgttgctgtt gttgtttgtt tgttttaatt 187260
gaaaaggggt ctcactctgt cacccaggct gaagtgcagt ggcgctctgt gggctcactg 187320
caacctctgc ctcccaggct gaagtgatcc tccaacctca gcctcctggg tagcagggac 187380
cacaagcacg tgccaccaca cctggcaatt ttttttttt ttttttgtat ttttggtgaa 187440
gacgaggtct tgccatgttg ctcaggctgg tctcaaactc ctgagctcaa gtgatttgcc 187500
tgccttgtcc tcccaaagtg ctgggattac aggcgtgagc caccgtgcca ggccggctct 187560
tgtttttctc ttccccctac accccaaata aacacagagc tttattcctg cctcagtcaa 187620
attgctgctt caaggccgca gtttggacac tatgtttttt agggtgtggt ttttttttt 187680
ttttttttta gacagagttt cgctcttgtt gcccaggctg cagtgcaatg gcacaatctt 187740
ggctcactgc aacctctacc tcccgggttc aagtgattct cctgcctcag ccttccaagt 187800
agctgggatt acaggcatgt gctaccatgc ccggctaatt ttgtcttttt aatagagatg 187860
ggatttctcc atgttagtca ggctggtctc aaactcctga cctcagatga tccgcccacc 187920
tcggcctccc aacctgctgg gaatataggc ataagccacc aaactcaact tataatttat 187980
gattaaggct gcagtgcaat ggcgcgatct tggctccctg caacctctgc ctcccaggtt 188040
caagtgattc tcctgcctca gccttccaag tagctggaaa tataggcaca cgccaccacg 188100
cctggctaat tttgtatttt taataaagat agcatttaat tatgttgtcc aggctagtcc 188160
cagactcctg acctcaggtg atccacccac ctcggcctcc caaagtgttg ggattatagg 188220
tgtgaacccc tacagctgac ccagacacca tgtttttatg gctggatttt gtctttgctc 188280
tggttgcggt cttaggcacc cttataaata gagctttgaa gagaacatta ccaatgtatt 188340
tttaatgagg tcatgttata aaattgtcgc ataggacttc tcaagaaaag acagcctctt 188400
ccttgcaaga tactttcttt tgcaaagatt gagatcattc cacaacaata gacctctgtt 188460
cattgcttcc ttcttatgca aaagtggccg tccctcccat cagaaggacc cccgctggca 188520
ctctgtcagg tagacagaag catggataga aggctggtgg tgagctccag gtgccttccc 188580
tattgtctct tctctcctat aacctcgtat aaccttcctg ggttttcctg ggtgcatgtt 188640
tttttgttgt cattggtgtt ttgacagctg gctgtccagg caaggctgct gtgtttgagc 188700
agaggtttgc tgagttgagc aggggtgtgg ctgcagggcc tagcctggcc tcccaggagc 188760
ccccgctccc cgtgtgccca ggtcataccc aaacaggagc attccttatg ctggtcctgg 188820
acagcgtttc tattaaaggg ttctttgtgt taggaatgtt cagcagagcg ccatgagccg 188880
ggtgagagtg gaatgagtgg tttacccagg gcacctctgg accctgggag tcacagctgt 188940
ggaattttac tggagttttc actgcagtgc agcccagggt aggacacaga gggcttccac 189000
tcccttggag catgctaatc tttccaaaac actcattcgt gggccctcat agaagctcct 189060

agggcattac caagaaatag cagtccttga tcatatccag tgaattctga aacagtgaag 189120
gaatttagat ctcatgtgtc catgttgctg agggcgtcct gggcacagag cctgctcgca 189180
tcaggccaga ttgtttggag tattgccaac tggccttttt tctggagaag aaagtactga 189240
cgctacgaag acttcagtgt tctcctgcag gggactgcag gggactgcag gggaagggag 189300
gattggcctg tcacttgcca tctctcattt ctgcgatgct acagagaggg aaggggaggc 189360
atacatatgt cagaatctaa attacagcat gtggaaagac ctgccctcgg ggtcagagca 189420
caccaggctg gggaggacct agtttaaagg gatagaagag acattacttt agctccttct 189480
cttcaggggc tccataatgg ttttaaactg ttctttaaaa tcgaagtttt tctaatctac 189540
ttttgactta tgtattaacc aagaacctct tgtaaatctt aagactatat agttgtcaaa 189600
gacaggcaac ttgaggttga gtctgttgct aactaactta ctgacttcgc acaaatcact 189660
ttgtctttgg gaacctcgct cccctatcag taaaacagag atgattgatt gattcaaaag 189720
cattcactga gcacccactc agctgcgagg cactgcctgc atacttggga tatgtcaggg 189780
agtgagagag gcaaggatcc ctgtcaacat ggagacttca ttccagcaga ggagacacac 189840
aggaatgagt gaatggaata agcaaatagg gtatgtactg taggagcaaa caagggatag 189900
aaacatggga gatgggtcag gagtgtggac tcgagtccac agcgctaaac tgtcccattg 189960
agaaggtgaa tgagttaaaa gagatcagga agttggccaa gtagatgcag gggaaaagtg 190020
ttccctggag agggagctgg ccagctgcat gcacaaggcc tgatggacgg ctgagccagt 190080
ggggaagagg aggcagcagc acagtcggga tgaactggca ggtgggcttc tgctccagaa 190140
gagatgcggg agccctgcag gtttgagtga agagtgcatg gtccaacacg gtcttcagag 190200
catcgttcca gctgctgcag gtttgagtga agagtgcact gtccaacaag gtcttcagag 190260
catcattcca tctgctgcag gtttgagtga agagtgcact gtccaacaag gtcttcagag 190320
catcattccg gctrctgcac agggaagagc atcaggggca agagttgatg cagacagtaa 190380
tagatggtaa tagagtcaga tacaattggg caggcaaygg ccctttacag atgacgaagc 190440
atcagaaaag ttagggtgca accatttgtt ttcagtttac aaaaagggaa gacgattaat 190500
ccccaaaaag gagcctgtga gagtcagatg aagaaattaa gaaatgaata atatgggtca 190560
catgagacag tctcttttctt tttattcatt tatttatttt tacaaaaaag tatgtttctg 190620
tgtccttcag cacagtttgc aggagcattt agagcacacc cgtggagtgg ccctttttatg 190680
cttgccaagc atgctgaaca ccgtaagcca cgtgtgacac atcttccatg gacatgaaag 190740
atatgttgat cattttattg ggctccagtc tcagctctgc cacgaactgg cactgtgcct 190800
tggaccaagt cacttcatcc ctttgggttt gcgtttgctc ccctggaagg taggggaggg 190860
gtgcagtgag ctctggcgtt cttcttagcc tctgctgcag ctgcatgagt gggtctatgg 190920
cacagccccc tgcctgcatc atggcaggtt atacacagta aagagatgaa aggaattttt 190980
ctgctaaggg aagtagcccc atctgtcagg atagttggct ccattgtgtc taacgtaggt 191040
atcttataag cctgtacaca tggcagccaa ggggacctgg ccgccagagc cgtaggagat 191100
gacccagcac aatgggctgg gcagtaagga agccagactc tggagccagc gtggaggtgc 191160
aggagctcgt gagtatgagg gcatgatgag gggtgcacag aggaacccct gggctaacag 191220
gggcccagga gacagtatta cggcattggg ctttgtattg ccggagacca gcacagatcc 191280
cacaatgcaa cgatgccaaa aaacggtaga actgaaaacc ccagccagat caacgcgaga 191340
ataaatctct tttctgctga aattgatagc ctcctaaaat gctaagacac atgcagygga 191400
gaataatcat tattgaccat gaaatagcta agaaccagct gagaaaatac agaaggacac 191460
acagtaagaa tgaatgagaa aactcttgca tagaggatac ggtcagagtt agcaaccagt 191520
tgcttcttca tgtaaattaa atcagcggag aatctaaaac catcccgtag accacattta 191580
gagggtagga aggatgcaat ggggcaaggt gggcaggaga tgggcttagc atccaagcag 191640
gctggactca cagccctctg cctggtgtgt gatctcagca cttcttgtac cttatctgag 191700
cctcaatgaa ggtaataaaa tcacctgcct ataagcctgc agtgagaatt agaggagcaa 191760
atggatgagc ctcagtcctg tgtggggtct ggctgctcac aaggcaccat ggacgccgtc 191820
tttaccatca tcactgtcga cccggagcca atggtgaaag caggacacag gcaagcccca 191880
gcgtttccca ccattgtctt atttttcgg cttcaggaag acattagact tctaggaaga 191940
gattccttaa agccaggact agaaggtaga ctccagattt tggctacaag tggcaaatat 192000
gtcttgtaag atgaatttta tgtacttgtg ccaagtgcca ttggaaatac cgaagactgt 192060
gcaaaaataa aagacaacaa acagccccag gaacccggag ccctctccca gcccagaaca 192120
ttcaccagct cggccaagag ttctgctggg ttttctctgg gggctggtgc tgctgtggac 192180
acgacaaccc ggaacacgga gggagggctc agcgctagga agggagaggg aatgaagagg 192240
agtttccctc tctttgctaa tttcttcgtc tctgggaaca tttccttcaa cagagtcctg 192300
cttttctcat cctcacacct cactgcgccc ctcctgaacc cactcctttc tgaatatggt 192360
ctactgtcct tccgtgaccc acatcacctt ggtcctctcc ctcataagca catcctaggt 192420
gggcctgccc ttcacttacc catctcctta gaagaaacgt gagctctcca aagggaaggg 192480
cagaaccctg cttgttggtc tttctgcccc cagcacttga cctagagcct tgcactgagg 192540
acgtgccact catgtctgct gaataaacag ccacatttcc agatgacgat gtccttttcc 192600
agccaacatc agctcagcgg gccttcacgt atttagttat acttgtgccc ccgctcaaca 192660
gggtgaggat gctcctggac acagaaatta gctctgaggc aggaaggagg aaaggggatg 192720
cttctgggag gcaaaggcgg tcaatcagag tgagcaccag agactccgtg tacctgggaa 192780
atacgtgggt tcccacacca gccttgggga gccagggtgg ggaagagggt ctgcagagca 192840

agtttaggat gcagcacatg ccaagctttt cagagtctca cagtcaggaa cagaactcat 192900
gcagggaggg gagggattgg aaagtaggag gcaaagcaga agccccgaac ccaaagacag 192960
agccggcgac cggccagagt gcagctctga gcctcagaca tgaggggaga agaaggggat 193020
ggggtggggg gcggtcgtga ggaatgtcgt tgtccaggct ccacccggcc caccagctcc 193080
gcagaggaag gagtgggctg ggagaggcac acaccagaac agctctcctc ggggcaaagc 193140
aggctttctt cccgaacacc caaggctttc caaaaggtaa acaccatttc ccccaagcga 193200
ccccaatgtt tgctgaagca aaacctctcg tgtgagccgg cgggcggctt cacgacaggc 193260
gtgagaaggc catggccctg tgtgggtgag gaagcgcagt gcggctcccc cctgcgtggt 193320
gggactaaga agagcccccc gccacccgaa aggcgcccta acacttcaga gagcggatgg 193380
ctgccgaggg tggccaggct ggagctgcgg cttcccaccc gatgcattgc agaatgtaac 193440
tttccaaaat gcattgctct catctcagct cagcgttaaa acacatgtgt gcacacacgc 193500
acatgcagcc ccgctgagct gggtggtgaa aagaccctaa ttagttctga ttccttaagg 193560
catgtatttt aaaaagcgtg aaacctattg agatgctact tcctagcgcg aatacggggc 193620
tcttaaaagt cctgataaaa gtgaaaatcc gaggcgcgcc tgggaagtgg gaatgttccc 193680
tccaactcag gcttccacgg tcatgagtag gaagtcctct tcctaatctc agtatcttaa 193740
aaagaagcct tgatgttgtt acgtgattac ctaaaaggaa tgccttcctc cgcggaccgg 193800
aaggatattt ttaaaggaat gtgaagcttg tgacaggaat tatcgatacc tttggaattt 193860
ttttttccaa gtgactcagg cttacttgaa gccattacct cggagttagt cagggactgc 193920
atgacgccag gccccaactg tttaaagcag agcgcggctt agtgaaagaa tgaaaaaacc 193980
gaggatgttc tttgtccatt attctcaccg tgatgaatga tgcttgtttt cctctccact 194040
ttaattagaa tgtttctaca tttgccaaag aaaatgttgg aatggagaca aaaacctgaa 194100
attataggaa cagggcttga tgtaatagct tatttgtaaa ggaaacacaa cttgtttggc 194160
attttattga aacaggaagt tcagaagctt agtacacaca agtacaacaa attctcaggt 194220
gcttgttgag tcatctgttg ttggaaatag tctcctggta gttttcccct tgatttactt 194280
tttatcttca ttttgttttt ttgaaagtag tgagggtagg aagttacaga gagattcaat 194340
tagagattat gtgtattttt aaaaatcagc tatcaagatt aaataaagca agcgggaatt 194400
ctctccttgc tcccatgtac caatttttgt aattatgtac aagatgaggg aaaccaaaga 194460
aaaacaataa cttgcttcaa tgcaattact aattcaaaag taaccattac tctggggaat 194520
tgtattagag attaacaaag aggaaaagta ctgtggtttt ctttctctat gttctatttg 194580
ctaggaagcg gtcaataaag taacctttcc cccacaggag ctggttaata gttcgcttca 194640
tgctaaataa aagttacaga aatatctgga gctgagttgc tggagacaca gaaatcttca 194700
ggttggaatt tcttgccctt ttccaaagga ttaggccagg acattgctgt caaatctgca 194760
aaacctactc atcctggcaa gagtgcggta tttttaggac tcactagtgt gctacttcta 194820
atagtgctta gtcagggacc cccaggggag tgcaagggag agagggtccc cagcagggac 194880
gccagacctt ctctagctgg ccgtgggtgc tggcctggcc acctgtagcc ctcagcgcac 194940
aggtggaggt gtaactggta ttcctgtggg agtgacagtg tccatctttg acatttaaga 195000
gcctgctcct tcagatacat ttaccattgc caccattggg gattggggca gtactggcca 195060
cccttggcgg cacatctcca gcttacagca gagtctgagt gtctctagca tacctctgac 195120
tgaggcasgt taggcttgtg acatcacatc ttcctaggtg gggcagagac tttacaatac 195180
atgtgacaag agaaaaacct tacagctttg tattgaaaga tttcttaagt ttttagttta 195240
ttgactaaat aacactgaac aaaatgattc tactatgaaa cgaaaggatt ggacctctgt 195300
gagggttgtg gcaatgtttc aatagctgag caacgcagga ggcacacagg ccatcgttgg 195360
gggcaggttg gaggccttca gttcctttac agctatgggc tcccatcaag ggtgagtgca 195420
ttgaggagac attgcctaga actactggac agacatctca cccaggagac gggagcatgg 195480
tactcaacac acttccatgc accgttcaga atcgctaaac acagcagtgc agaggcagat 195540
gacaagggcc attacggggt caccaaggga ggaaataggg actggagccc ccaggaagga 195600
gagctgagtc tccctgtggg ctgggggctg gctttgtggc cctgcagcca ccacctggag 195660
atgagagacc tgtccctagg cctccctgca gccaccacct ggagacagag tccctagacc 195720
tcccagctgt gcccacctgg gcagctgcac tttccagagg attattcctg cagcttccac 195780
cctcacatct ctcagctgtc tttgcaggtg catctctgga aaacagttct catcagggca 195840
ccctgtgctt cccagtttct agtcatttcc cttctctgaa ggttctagtt cagactcttg 195900
agcaaagcct tcaagacctc tccttaaact gccctcctcc tcttccgtcc agccacctgg 195960
ttgcctcctg gctcttcctg ccctaatacc ggctgcccgt acgggactgc tcacctcctg 196020
cagggagccg gacgtctgtg gcgatctccc tcccgccatg acacccccta cctgtcctcc 196080
atcatatggg acacacacac acacacacac acacacacac ccctacgcac acccacaccc 196140
cacatgcaca tcatacatac atgcccacca gaaatacaca caccatacac accacccacc 196200
cacatgcaca ccatacatac acatacacac aacacagaca ttaaatacac atgccactac 196260
acacagtgca taccacacac aacacacacc acacacacac acccaatcac atcacatata 196320
cccacaccac acacacacac acccaatcac ataccacata tacccacacc acacacacac 196380
aagcctttcc taattatcta aaggagaagc ttttctggaa agcattcccc agagcttcta 196440
gagaaattag tgtcaccctc ttttatggtt tcatagtaat gtttttatat caccagtata 196500
aatactatca tataaaaagg gtaatcagtg taccatagta attaattctt taagtatgtc 196560
tcttctgcta gatgatgagc ttcctgaatg caggctctga agaatttttt catagtttta 196620

aatccactgc atggaataga gaaggctctc cataaacttc ctgagtttaa atggaatcgg 196680
attggaaggc agtagcaagg cacaaagtgc agtgagagcc aagctcagga aaaccagtgt 196740
ccttgagcag aaagacttag gaagggtgct cgctagcgag gagggaggca acaaggggcc 196800
agcccgtggg gagccttaag caccaagagc agggcggtgc acactttgtc tggcacgggc 196860
tggagcagga gagggaccgt ccttgcattc tgtgcggatt tctatggcaa tgacatggag 196920
ggaaatgaag gtaggatcaa gagtcccact gggaagtggc ctggcaacca gaggtgtccg 196980
caggacacct gagcctcagc agtgtctgtg aggataggag ggaaagccag accccagcct 197040
ctctggggag aatctggatg catgcgggag gaatggatgg aagggagggt gtggggctga 197100
gtggcggcgg ctgggctgtg ctctcccact cacagagcct tccccaaagc ggggaaggct 197160
gcttgccttt tggttcattt cctttcttta atacacagca aattcctggt cacccttttgt 197220
tgttggctgg ttgggtttgt cgctttcctt gttgtttaca agctccaggt atttgtgaca 197280
gatcttatca tctccttccc tcttagtcac ctcttggccc aactctgcat attttacctt 197340
tttaactctg ctcctgttct gacctcccca ctctcggaag catatttgct tggtgttttc 197400
agattttatt tcatttttggc tatttaaaga gatgcaataa actaaatatg gcctggcaag 197460
tctggtctta aaatagaaaa tatatatata tgtatatttg tgtgtgcatg tgtgtctgtg 197520
tacataggcg catgtgtgtg cccttgagtg tgcatccgtg tgtgtgtgtg tggccggtgt 197580
actataaacc cagggcatca gtctcctgac gtcattgctt gcacttttttg ccattctccc 197640
ccaaacacta gttttcagcc tgtattttct cagtttcccc aaaaatgatt ttttaagaaa 197700
agtcaaatca gaaagtgatc agcctctacc gccggactct gcttcagtat ccatccatgt 197760
ctctgaggtc ttggggctca taggaatgtg cttattttca tagtcccatt aacatgaata 197820
gtttcagaag ggccagctca gttttgtctt cagttttctc actggtgatt gtgcaggggt 197880
ggaatggcaa tggaatgcat aggggcatga gtgaactttt cgggatgacg gaagtactct 197940
atattttgat tgaggtgtta ttaattcaat gtgtcaaaat catcaaaatt tttacatttc 198000
atcattctta aattatacct cagtaaagtt gattttaaaa gttaaacaca tacccttttgc 198060
tcgaaaatga tcctgtagag cgtttatgcc tttatatgaa tttagctaat gcattctctc 198120
cccagggcca tttgcatttt aggatataac tgatgatgtg gaaggtacta gcaaggaagt 198180
atgggatggg aatctgggga tggaagtacc ttcctgcttt cagtaagtta cataggcact 198240
ccttattcat aaggctgagc ttggtttcag ataaataatc agaaagtagg ttgtgcaagg 198300
ttttaagaag aggatccaaa ctgggactta gtaacgaact ctgaaactgc cacttgcatt 198360
ctctgaactt cacatcaagt caatactctg tatgctacaa ttccatctta cattaaaaag 198420
caggtctact aagggacccg attcccaaga aataaatgtg ctttttacaa tgcttgattt 198480
gcaagtcagt ttcaaagata atttggtgaa gatatcagag ttatttttac aagattaaaa 198540
atcagtattc aacaaattat tttattcact ttgacttttt ttttttttta acctgtctgt 198600
gacatatgtc tcctttgatc cgcacacaca ccctggccag taggaaacag gcacactctg 198660
ctggtggcag agggatgggg actggagcct gatcttggac cttccctgtc tcatctagct 198720
cagcccccat gctgtcatag gccgcagcca agtggccttc cacagcccct ccatggagcc 198780
atcgcagaca cagcttctcc acggagccct gttctcagcc ctggaggccg gcaatgtgct 198840
tcacccactg cctgccacat tccagccaac agaagaactt ttgaccgaga agtagaaact 198900
aggtgattca gatcagatct ctgttgtaga ctccactacc ctaatgatga atttttaaaa 198960
ttaaacattc cctaacaaac ctccaagact ctttgcttgg gtcggtcaaa atacagtgga 199020
atgtgagagc acatgtcaga attctccagc ctacgtttgc tgttgttgtt gttttgagac 199080
ggagtctcac tctatcgccc aggctggagt gcagtggcgc aaactcggct cactgcaatc 199140
tctgcctttc aggctcgagt gattctcctg cctcagcctc ctaagtagct aggactatac 199200
gtgcgtgcca ccacgcctgg ctagtttttg tatttgtagt agagacaggg tttcaccatg 199260
ttggccagac tggtctcgaa ttcctgacct caggtgatct gcctgcctcg gcctcccaaa 199320
gtgctgggat tacaggtgtg agacaccaca tccagcccag cctactttta tactatgaac 199380
aaaacttctt agaattacca acttaagtac aatagaagct tttgaaatta gctgggggga 199440
aattgagtct ctaagtaagg aggagtaaga gcaagaagat cagaaggaac cacagaatca 199500
aacactttca aaaggaaaga aaattaggaa attgttcggt gccatccctt catttcagag 199560
gggaagaact aaggactaga gaagtcaggt caccccgaca ggaccctatg tccctccttg 199620
tcgcctgacc tctccctgtg agtctcagtg gtcctggtcc cacagcaggt gcttggggac 199680
ccagaaagag gccaggtctc ctgacaccca gccccgctct tgttgggtcc ctgaatctgg 199740
aatggttact catgttgggg gaattttata ttctttttc caaaagttga tatccagcta 199800
gaatctgtcc ttcctgagag cttgtcactg ccctttctct cctccctgcc tgtactcctg 199860
ttcgcttggg actcacactc cttgcaaaaa agcttgtttc acccaggggt gagttttgta 199920
actagagcag ggagtccttg cctttcattc caatgcattc cccaaaagca gaaaagtgtt 199980
atgcgatggg agtttgcatt ttggaccaaa gactccgcag caaataaatc atggaaacga 200040
acaatatgtc cttaaaccaa gatgtaactg taaacctcta ctgtcttatg aaataacaat 200100
actgtgcttt gagtagccag accacatagt agctggactc tagactctaa gcagggatga 200160
agtcagtggc tgctgatctg ggccttcccc agaaggatgc caagagatca agttttgttt 200220
ttaagttctg tgaatcacag acattatttt tgtaatcttt ttttttatga cacagagtct 200280
cactctgtca cccaggctgg agtgcagtgg cacgatctca gctcactgca acctccacct 200340
cccaggttca agcaattctc gtgcctcaga ctcccaagta gctgggatta caggtgtttg 200400

ccaccatgcc caactaattt ttgtattttt agtaaagatg ggtttcacca tgttggccag 200460
gctggtctcg aatgcctgac ctcaagtgat ctacccccct tggcccccca gaatgctggg 200520
attacaggca tgagccacca tgcctggctt tgtaaaaaat ttttaaagcc aatttgcttg 200580
tttaaaaaac tgaatccaca ctggtaagtt ttgttttaat aaaaaaattg tgagtaagtt 200640
gtaaagcttt tgataagttc agtggctcct gtaggcagac aataaattgc taagtcccaa 200700
agtgttgcaa gattctggag agtactttgt tcatactttg aagaatatgc ctgattataa 200760
ggcaacacaa attactgaag ccttgaaatg atgaggttgt ttccatttac tcgcacataa 200820
aataatatat ctaaaacatc tagcaactct caaaagaaga gagtaaaaag cttttgagaa 200880
atcaaataca attcattcca attcaacttg aaaattccca acagtccgtg ttgcatttta 200940
tacatcttga accaaaccat ggctttgagt aaaggcttca tttaaaaacc taacctatat 201000
atggtgggtg ttcatgttct attaaagcaa ggtccctgtc ctagttggag ggaacttccc 201060
taggttcggc agcataaacc agtgcctgtc gaccagggag tgtcaggagg atgtgctgct 201120
tcctgccccc tcccgcacag ggagcaaggc tgtgctgaat ggagatattc tagtaaggag 201180
gagagtgtat gtgagaaggt gtatgtgaga aggtgtggca tccacaacaa aactaataaa 201240
gcatcagcaa ccttaggtga tgcggtttgg ctatgtcccc acccaaatct catcttgagt 201300
tcccacatgt tgtgggaggt aattgaatca cagggacagg tctttctcat gctgttctcg 201360
tgatagtgaa taagtgtcat aagagctgat ggtttcataa gggggagttt ccctgcacaa 201420
gctctcttct cttgtttgcc accatgtgag atgtgccttt caccttccac tatgagtgtg 201480
aggcctcccc agccacatgg aactgtaagt ccattaaacc tctttctttt gtaaattgcc 201540
cagtcttggg tatgtcttta tcagcagtat gaaaacagac taatgcattt ggaaaccaag 201600
aggctgatgg tgttcaggac acactgtccc catttatagc accttggcat ttcagaaaat 201660
cgcaaaagca ggaaggcccc tctcactttc ccctccttgc ccttctcccc tggggcaggt 201720
tataagatcc tcatttggga gagtctttcc caatacttgg aggaaaggaa catccttgtc 201780
tctgaagaca cagagcacag agaagaatca gaacaaacag gcctttctca gtgaccccag 201840
tttatcacca ttagctcact cccagtttgt ctaatcacct cctccaccac tatccactct 201900
tcatcaaacc taagtacaaa atacccaagt ttgcctgttt ctgtgggtct tcctttcctt 201960
gtgataactc ctgagtcaca tgaaacacat actaaatatg tgtgcctgtt ttcctcttgt 202020
tactctttag ttacagggaa gggccccagc catgaaccta gcaatgggtg aggaaagaaa 202080
tctttccttc cctactgata tggtttggct gtgtccctac tcaaatctca tcttgaattg 202140
tagctccctc aattcccatg tgttatggga gggaaccagt gggagataat cgaatcatgg 202200
gggcagtttc cccccataca gttctcatgg tagtgaataa gtctcatgag atctgatggt 202260
gaataagggg aaatgccttt cacttgcttc ccatttttct ctcttgtctg ctgccatgta 202320
agacatgctg tccaccttct gccgtgattg tgaggcctcc ccaggcaggt ggaactgtga 202380
gaccattaaa cttctttctc tttataaagt atccagtctt gggtatgtct atatcagcag 202440
catgaaaacg gactaataca cctaccaggc ccggatttgt ttggcaataa agtgatccat 202500
tcacgcccaa gaagtgggtg gagctgggaa aggccagacc aaccatttgg aatagtgttt 202560
tttgatccac ccccaggagg tgaggattgg caggggctga ggggagtgct cacctccagc 202620
aaggtgagct ggagcccaca gcaggactcc agcctcagca gaggaactgg agagcaaacc 202680
aggaaaggca gacagagctg actcacgtgc gagggtggga gaggtcgcac ggcctgcccg 202740
gaccctgatg agctgagcac agtgaaaaca atgccaggcc tcacctgccc gtgcttaccg 202800
gctggtggca gggggggctga gcaggtgttg aggtgttcac aggtgagtag gagaggaaag 202860
gcagacgtcg gcctaaaggc aatcgcaagg agaaatgcgt tgagaattgt agcactgtat 202920
ccatcaaaaa ggaagctcat ctttcactgg gtgtctttct aattgttaga cttgacactg 202980
catttgctgc cctgatttct tgtcctaacc ttcaagcttg ttagaacagg gactcaggga 203040
ctctgttttc ttctcctgtg ctcagtgcag ggcagcagga ctcacttgct aagtgctcac 203100
tgacagatgt aagattattg ttagagatat ggacccgctt gctcttctga gcttccgtga 203160
ttctcattcg gtcctttgct gtcattagaa tcgtctgggg agaattttgt cactcctgct 203220
actctgacca aacctcgtat acttcaatca gaatgctcgg agttggggct gcagcaactg 203280
gaattgtttc aaactccccg ggtgactgcc ctagcagtca agtttgagaa ccacgggcat 203340
ggtaaaatct tttctcagcc tgagcagccc attagcttca cctagggagc tttaacaatc 203400
actaatgcct aggcctcacc accctccatc ccgtgttctg acttaattag cgtggggtgg 203460
ggcccctaaa acaacattct aacagcttcc caggcgatga gaatgcacag ctaggatgag 203520
cttctcctct gaagcatgaa gacccacaga atactgcaga gttgctgggg gtggccctgc 203580
ccaaattctc gcctaaaacc ccaactttca atgacattgt ggacctgctt tcgtgttatt 203640
ataaggttta caaatttcta tgccacctat cagaccattt tttaaggatg aaatcaaagt 203700
ttctataagt tgtatagttc tttccctgtg cattttatcg taatattgaa aaacgacagt 203760
gaaaagcaac caaggcatct cggcagcatg ctgctgacta gttcacgcag ttaccaccaa 203820
agcgcatgga cgggacccag agcatsagcg tgtgcccact atcggggaca gaaacctacc 203880
gcgttcgagt tttgacatat ttctcgcagt tgttgaaaac tatgaggcat gaaatccaga 203940
tttatgactt tttaaaaagt tatttgtgga ttcccaagac gattatgttc ccatcactta 204000
tgtagcctta aaagaaaaaa acctcaaatg atgctttaaa aaaatccaag tttggcgctc 204060
attgagttcc agtgtcagtt gtctgaatcg ccttcagcga aagtcagggg gaaaaaatac 204120
attccgcctt cctttaactg ctagttcgtc atggagaaca gaaagtccca tttgcatgtg 204180

gcttttggaa aagctaagcc gggagcgatt atcctgatgc gcttttactt tttgcataaa 204240
ataagaattt gaggaggatg tcccgggaga gtgagccact tctcatttcc caggcctcgc 204300
ctgccatgct ctttgacaac atcatagatt ttatttttgc cgggaatctc attatcaaag 204360
caatgccccc cgcccccccc ccccacacac agactgccag gtaaaccaca gagggtgagg 204420
ggggtgcagg tcatggttgc cttattacac accctcctct gccatcacct ccttttttgt 204480
ctggataagt tctttggcag ttctctcaac ttttatttct gaaacatcct gaaacatctc 204540
agtattaaaa gcaaggccga ttatataaac gatactccca ggcctgacaa cacatggttt 204600
tgcctgaggc ctttactgcc aagagccgta aggaccctct aagtcatgtt cgctatttt 204660
actggccttg agagtctcct tgctttgaca tcctcttgtc tccattgtca gactgttaaa 204720
tgctcatgct tctggttctc ttaaatagat gcagatgtgt ggggctgggt tgccactgag 204780
ccctcttctc ttttgcaaga gctgggatgc agacagaagg cggtttggaa aacacgagcc 204840
accttgattt tagacaaact ctaagttaca atcaggtgtc ttcatttatg acatttaact 204900
tttacttaac ctaatcaagc catgttgttg gctactgatt agaatatcct tttataactt 204960
accttaaatc tcactacttg ttccaaccat cccaaagtct ggcgtcaact gtcattgcat 205020
gctgctcttt tcagcctttc tagttcgact cttagcaaaa gccataatct tcctccagtc 205080
tgtttccttt ctgcagtgac aaaattgccc agggaaagga aaaagaacag catctatctt 205140
ctttcttttt agctccctgg tttaaggctt tcttttcccc catgatgaaa aactataatc 205200
attctgctta gaaagtacag acccctaagc ccacttccaa aagaaggatg cattttcaag 205260
tctgttatct ttactttccc agagcctggg ggtctcccag gccagaagtt gacagaactg 205320
tcttcataca ctcgagacaa cttcatgccc atttccttaa aactaagaac ataagacgct 205380
gatttttctt ccagaaaaaa aaaaaccttt cttgttcttt caagaactgt ttcacggaca 205440
gtgtttcata ttacaaaatt gaaacttggg acttttgaac tgcaaattta gcagaaaatg 205500
aatccatgcg cttgtggctt tgcttgtcac ctctactcag atgtctccca gacccctctc 205560
cagctgcaag ctgcaggcag aactgttcct ctaaaagaaa acaaactcct gtttttccta 205620
ctactgctac tgcttctact gttgctacac acacacatac acacacactc tctcacacac 205680
acactcacac acacacacac acactcagaa aacacttctg acaccaaatg tatgggtttt 205740
tttcatgcca aacaattctg cagttcactg cagacaccag ctgagtgtcc tacaatccaa 205800
ttgtggcacc gcctgcctgg agttagcagg tgaaggactc agccccgcaa gcctgccccc 205860
ctacccatgc caattgcttg tcccagatcc ccgttctaac tgaccagcgg taaatcaggg 205920
gttgccacaa ccccctcctg ggatttgtaa cttgctacag cagctcacaa aactcagaga 205980
aacacttaac attgaccaat tcatcacaaa cgatattttg aaaggatgtg aatgaacagc 206040
cagagaagag atgcacaggg cccggggccg gggagcaggg catacggagc tgccatgccc 206100
tctcaggggg catcacctcc tgcaccaggg tgtgttcaac cccaaagctc ctgaaccctt 206160
taacgtcagg attttttttt attttttttt aaagacatag tctcactctg tctcccaggc 206220
tggagtgcag tggcgccatc tcagctccct gcaagctccg cctcccgggt tctcgccatt 206280
ctcctgcctc agcctcccca gtagctggga ctacaggcgt ccgccaccac gcccggctaa 206340
ttttttgtat ttttagcgga gacggggttt caccgtgtta gccaggatgg tctcggtctc 206400
ctgacctcgt gatccaccca cctcggcctc ccaaagtgct gggattacag gcgtgagcca 206460
ccgcgcccgg cctaacgtcg ggatttttaa ggagcttcat tacataggca ggactgatga 206520
aatcattggc cattgagtga accccagacc ttgcgggggt ggggctgaaa gtttcaaccc 206580
tccaaagatt gggcacgttc ctctggcact cggcccccag cctccaggag ccacctcatt 206640
agcatacacg caggtagggt tggaaagggc ttgtgataaa tgatgaagga cgttcttctg 206700
catcgctcgg ggaattccaa gggtttaggg gctcactgcc aggaacccgg ggcagaaacc 206760
aaatacatat ttctcgttat agcacagtgt caccccctca ctctgcctaa tttggtgact 206820
agctgcccca tcacattctg cctatttaag ccaagccccc cttccccaag gccaacctcc 206880
tctcctccac agccagccca cttcccgggc gtgataactc ttctgcctca gctggagagt 206940
tgttctgagg ctttcatcct tctccacgtg ccgcctggca gtgctgctgc ctgtcttttg 207000
agggctaccc ctttctccat tacctctgcg acctggctag tccacatcct ccccgacccg 207060
tgctcttcag caccggtgcc tgccccgctc agtgcatgtc ctcatccctg cagcctccac 207120
cctgggcttc ctgaccccca ctgcgtccgg caccgctggt tgcgggcctg ctccggctct 207180
ctctgcccag ctggctggcc tgcctctgtt ccgacctccc ctgcctggcc tggtgttctg 207240
ggcgcctcct ccgctcacat cgccgcttca cctgcttttg ctatctgcac tttccatgtc 207300
ctgctccttc tcccagctgg tggtgcctct gagaagagga ctgagaaccg cctgtgaacc 207360
ccgcaatttc gtgggtgtgg tggaagcaaa ggcagagcgt gtgagtttag tgggcgtgcg 207420
ccactctttc aagaagtttt gttacaaaaa gatgcaaagg aagtgaagag ggaaggggtt 207480
tgcaggttgg gagaaataac agcatttgtg ttgtttgttg ttgtgacggt tttgagccaa 207540
aacatgacaa acgggacaga aggaagacct gatggagcgt gtccttgaga aggcgagagg 207600
catggggttg gcctgctggg ggatcggcct tccatatggg ggttcctctc cagcagcctg 207660
gggttctgag gaaggcaggc ctgaagcagg tgccgggtgc cgggaagcag gagacatctc 207720
tgttactcca ctgtcctcag tggggagcca cggctgagcg tgagaaaggg cttataggct 207780
gaaggccagg cagacgggaa tggccaggca gaggagggga ggacgagccg ggtagaaaca 207840
gtggatagaa acacggaggg ccacacggcc aacggtcagg ggactggcac accagccaga 207900
ttcacccgcg gcgatgccgg tgcagagaag ctcggcatct gaatttaacc cgggttgtgg 207960

tttgactcag tctgacgtgg agagaagggc cagggagtca cgggggggtg gtgggctgtg 208020
tgctggttta ggggctggga catggagggg tgaaggcggg agtcagtcgc atccgctggg 208080
caggggcctg gggctgcaga caaggtggga ggtggcagct acggaggaag ctacaaggga 208140
ttctgcagtt ccccggggaa acaggagccc aagggaccgg ggggtgaggg ggttggaagg 208200
ggcacctgtg gatgttctga gacttccagg aagtgggaca ggatcagtga tggagataga 208260
gacagagtca tcagggccga gaggaatgac agtaacagcg aggttgaagt gggcaccccc 208320
gtctagcagc acggggtgtg gagctggctt gtggacggcc agggaacagg acgctttgag 208380
gtggcagcca ggggcaggga tgcttttgat cgccaaggga gaagacttga tgcagagttt 208440
caggagcctc catgacttcc ccatctgaag acctttttta ctttaatggg attgaagtga 208500
tcaccagaat agttaatggt gtgctccgtt cctatttctc tggtttttct aaggtccaca 208560
ggctgcagac atcgtttgta cttctccccg gtgccaaaga ccagttaatg ccgactttga 208620
tgggctcagt gcaggccaca ttgtcacgtg taactctaca ctgagaatta ttttagaagg 208680
ttagactcct aaaaatgttt tgtttttcca aatggtggcc tctgggtctg acttcacctc 208740
ttttgcaatg atcagcacta ggatatggtt ttggagacgg ttgtgcagag ccagggcttt 208800
caccaaagct tggccgctcg gacaggactc acgatggaag acggtcaggt gccccaggtt 208860
tcagatgcct gcctcctccc atgcgtggtg aggggcctgc ctcctttata gctttccgct 208920
gccaggctgg cgcctcctcc cctcaccccc atctcctcca gaggaagacc aacttaatca 208980
aatcttacca caactacgta ctgcctcctg gaaaaagcct gatttctcgc cccctcttgt 209040
ccctccctgc gtggaggcag gccctttgtc cagtgcccat gtggcttggt gggtggtctt 209100
tctaagttat cagaggacat tagcaaacac acacgtccat tggcctaacg cccaatctgc 209160
agccagcctt atgaataatc aacgtgactt gtctctgtag ttcaatgcct atatctgcct 209220
ctcagttgtt attgaagctg ggggcaaaaa agatggatta ttcattggaa acctcaaaac 209280
ctcgacagct gagctttctt acacatgcct gtgtggcccc cgtggtatct tagtgttcac 209340
ctccccattt gcacacagga agccagtcac attactggat tcctggtgag tttgactttt 209400
cattctgtct tgaatctccc tcccttcccc aaccccatac cccaccctac tccatccctt 209460
tttcttgggt cttcctgatc tcaacccctc catctgtcct ccacgttgtc tgcatagtga 209520
gcctcctaac acacggatcc ccccatggcc ttgtctgctc aggtttctaa ggtccccagt 209580
aaccacgctc acactgcgta acacgaacgg tctggtccac acctcatcac ttggcgtgca 209640
tgtgaatgtt ttagcaagtt agctcttgca attattgcct gccgatcccc tgggctgcat 209700
tcacacatgc cgtgagtctt cagacaccca ggtctcagga cctgaggggc tcctgtgtgc 209760
tttccgtgag gaactgtctt tctgctcacg actccatgtc acatgccacc atcaggaagt 209820
cctccctcaa tgccccaagc ctactcaggc tcccactttc ctgcccatga aatgtgtgta 209880
acttctaggg tgtcctgaga agcaaagacc atgtccctgc atttttgcat cctcagaact 209940
tagcctgata ctcacaatga aatgagttca cttaacgaca caacgaacga atgtgcaggt 210000
acttctgcag ggggtgatgt ggggatgcgt gcattgattc tgtggctcag ccctgagttg 210060
ggggcaggag gcaggtgctg ggaggaggat tttatgtctt aggaagcaca ggaaggcctt 210120
gccaggatcc aagaaaaaat ggaaagtaga ycaatgtaag cgttaaaaga acacatttta 210180
tcttttaaat gtgtgtacac agtacagttg actttttttgt atacaattct atgagtttaa 210240
acacacatat agattagcgt aaccactaat tataagattg tagggaactg gggaaaaaat 210300
gcatgcatta aggaatgata yggcatattt gggggacaga gaacaggctt gatgaggaca 210360
gagtctattt aaaagagaca gtgggcacsg caattggagg ggaaggcggg gcagggtttt 210420
agagaacccc tgagtgctgg gctacaggat tcagtaaagt tattgatgag attggctgca 210480
ttgtggattc tgaaatattt atttaatacc tcgaggaggg tgtgagtaga ttgtgctgat 210540
gatcgcataa ctctgactat actaagaacc actgagttgc acccagagct tgcattactg 210600
agcgctttac cagttaggaa ggtttcgcgt attccgtact ttaaatctaa ggtgacttga 210660
ctgtaaggcc tgcgagtatt tcctggacca ctcagaggaa gaatgctgtg aatgagaact 210720
acagccctgt aagacacgtc ctgtatcgtt gttgagatgg gaaagtgcat cttaagacgg 210780
ttagcaggcc gaggagcgac tttaaagggt gagctctgcc tagagggaaa agcgaatgca 210840
ctaattgaaa tccaacaccc tgggctggag taaatgaacc gtcagccacc catggggctt 210900
catttcttgg tgatggataa atagctggga ttccttgaag ctagaagcca tggggaaatt 210960
ctgttctgct tagctttgtc aacagtacag tctgccttaa ctgacttgga ggtaaataga 211020
ttcggagagt gtgagctaaa acccattaaa tcaggtgaag acacaaaggc aagcacagcc 211080
aatgtggttt aaggcaaagc taatgtcctt cggccttaac tgacggactt tcctagcagt 211140
cctcaccctc tgcaacccag ggctcctrgg aggagctcat ggcagagaaa gccttctggc 211200
ttctgccact gcctcctcaa ctacatgtat acatcagtgt atatgcatgg gtatgaaatg 211260
aacattttat gtcaccatta gcagaggaaa gctggaactc tttcaaaccc cacccaaaat 211320
tcactctgac tactgagcag tcctgttgtt tattttggag gccacttaac cctggagcag 211380
tccataagct ccacttaatc ccctcttctt tcatgatttc ttttaaagag acatcttggg 211440
ttctgtaggg gaacatttgt gcttcactgt aaaactccat ttgaggcctg ctcacggcct 211500
gccaccttat ctgcttgcag ccttcattgc ttgggagctg ttttacagct tcataagttg 211560
taaatagctg ctggcaatgc aaacgcgctt gtctgtgggc aggaaatgaa ttctgtctgg 211620
tagagggaat gcttcctacc ttgtaggaaa gccaatattt tttgtccatt agcaagttta 211680
tatcagtatt cctaatcatt aaatgtgttc ttcggattgt cctttgaacc agttatagca 211740

tttgagttaa gtaaaatgaa tacactgttg tttattttat acctgtatga aagttatggg 211800
ttttttggtg gggggggggt gttttttttg ttttttttt ttgttttttt tgaggtggaa 211860
tctcgttctg tcgcccaggc tggagtacag tggcgcaatc tcggctcact gcaagctcct 211920
cctcccaggt tcacaccatt cttctgcctc agcttcccaa aagttatgat ttttaaaaaa 211980
ttatctttta acatttttta gctagaaact tctgggtcaa tatataaata gatgagcctg 212040
gttatatctg aggttttcac tgaggtaaca acaaaaataa aacaacacga tgccaccgag 212100
ccatcgttcc ccaacttacg tctgtcccct ccacatgtcc tgcacacact cctgtttctg 212160
gggtgtgtgc atgtgtgtgt gtgtgtaaag gtttgcaatg aaattagaat cattggtttt 212220
tgttgggggt ggggagttgt attgttttga gacagggtct cgctctgtca cccacgctgg 212280
agtgaagggt cacaatcaca gttcactgca gcctcaactt cctgggctca agtgatcctc 212340
ccacctcagc ctcccaagta gcggaaacta taggcatgtg acaccatgcc gggcttgctt 212400
atctatgtct gtctgtctgt ctgtctatca tccatctatc tatctatcta tctaatctat 212460
ctatctatct atctatctat ctatctatct atctatctat ctatctttct atctatctag 212520
atggggtctc cctatgttgc ccaggctggt ctcaaactcc tgggcttaag caatccaact 212580
acctcagcct cccaaagtgc tgggattaca ggtgttagcc actttgccca gctgaagtta 212640
gagtttagag cacattgctg taaattgcga ttaccaaggg tattgaaaaa tccatgaaaa 212700
taataaacag caagttgact tcagaatttg tgcgtttgag gcttttcgcc ttgatctcca 212760
ggtaacacac aggctccttg gcgagagcca gtggtgatac aatgagaaca ccgcctgctg 212820
catctaatat ttgcagctta gaattcacag ctaacttttt aaaatgtacc agtgtggggg 212880
aaatggtgct ttatttgctg gataggaaaa ttggccaaga tcagaattct gaaggcagtg 212940
tcacagcaca aagaaactag ctactgaagt cacatcctaa acattcgaga ggttgatttc 213000
cttttctact gcattacaaa aaggtttatt tactgcttat ccatatagtg agatagagat 213060
tagatctcag tttttggtta agaacaagca ttatcataaa tgtgtgtgtg tgttgtgtgt 213120
gcattttaca ggatttttaa aaatacacag agaatttttc acagttgtta actctggtaa 213180
atggtgggga aggcaggggt gagaactgat ctattattca taatctcaat gatgaacaag 213240
ctatttccaa aaataggtgg attatttaaa attattatta ttaggatatt ttgggcttct 213300
agaaacaaaa acttaacaaa aaagtcactt aaagaattta ggggtctttt tttctgacat 213360
gaaaagaaca aaataaagga tgatttcagt ttggtccgtc agtgacttag aagtgttttt 213420
caggacccaa ggctttccgc cttcccactg ggccattttc agcgtgtccc gtggcctctg 213480
ggggcttcag tgatccaggc gtcacattag acatgacagt gtccagcaaa gagaagtatt 213540
tctgctttgc atctgtttat aacagtgaga aaaactcccc cagaatccca ccagcaattg 213600
attctcacgt tgcattggcc aggattgagg ccagctgtgc catgcttagc gcagtcattt 213660
gtattgcgat caccgtgatt agctcagacc catcctggga cttctccttg ggcttgaaga 213720
catggccagg tggagatcgg tgccccccag aagaagtctt tgttctgcca ataaagaaga 213780
cacagacaac agtgtctaac aggaaaagcc ccttttttact ttataccctt ccgtattgct 213840
tcaacaatca aatactttat tttattgttt gagacagagt cctgctgtgt cgcccaggct 213900
ggagcgcagt ggcgccatct ctgctcactg ccacctccac ctcccagatt caagcgattc 213960
tcctgcctca gcctcccgag tagctgggat tacaggcgcc taccaaaatg ctcggctagt 214020
ttttgtattt ttagtagaga tggggtttct ccatgttggt cagaccggtc tcgaactcct 214080
gacctcaagt gatccactca cctcagcctc cccaagtgct gggattacag gcgtgagcca 214140
ctgcgcccag cctttttttt ctttagatag agtgttgctc ttattgccca ggctggagtg 214200
cagtggcaca atctcagctc actgcaagct ccacctcccg ggttcacacc attctcctgc 214260
ctcagcctcc cgagtagctg ggactacagg tgcccaccac cacgcctggc taatcttttg 214320
tatttttagt agagacaggt ttcaccatgt tagccaggat gatcttgatc tcctgacctt 214380
gtgacctgcc cgcctcagcc tcgcaaagtg ctgggattac aggtgtgagc caccgtgccc 214440
ggccagatac tttcataatt aacttttga atgtatgtgt gtcctacttt aaaatgaaag 214500
atactctttc ttgattccat ttccatgcag cttggccccg tgatgctagg gaccatggct 214560
ttttcttgca gtgtgactca ccatttgcca aagcaaatct cttgccttgc atcagctcag 214620
tctctttgtc tgcaaattaa atcaatagcc ctttccactg cctatctcgc aggatatagt 214680
gccaaaaata ctcacaaagt caccatccag gaagaatcat ttgcccctgc tgccactgtc 214740
tcctgcaagg cacatgaaag ctgctgaggc tcggtattta ttatgctata aaattcaaca 214800
caaggggaga gaacaagcaa attccatgag catatataag tgtatcggat ctactccatt 214860
gatgctggag ctatattttc acagtaggat cctcttttgt taaatattac agtagtagga 214920
aaacctagca gaagaatagt tcactgtttc tctgattttg tgagtgatgt gggctgtgga 214980
atttactctt tgctgctctt cccccaacct gcaccctacc cctgcctccg aggtcagcct 215040
tgcctgctgc ccctgactga gaggaccccg acgtcacccc accccaggtt atactcctct 215100
gagaaggtcc cttcatccct tccccgaaat acatcccctc aaatctctaa tttgtgtgaa 215160
ccattaattt cagatattgt aggaaaaata agcagggaaa atacgcaaaa caaaacgtgg 215220
atggcacata acccatagca tctcgcaggg tgtgtacact gaagaagtct ttaccaaccc 215280
gtagttagga aaatgcgtgt tcagaataac tgggccttcc cgcggtcctc tgagtcaaac 215340
agatgaccac acattgccag aatgagaagc agagcagctt cacatccctg cttctgaaat 215400
gtttcccaac agctcattga aacaatctcg agacacctct ctcccccaaa cccagcgtgt 215460
ttcgggaatg gctctaggaa ttctactttt gcattgcctc actctccctt tccccgtcca 215520

aaccatggta ttggatttac agcatttctt acatcctata aaagtccttt tctgccaaga 215580
gcctggagcg cgctggattg aatgacgctc tcccagcaca gccggcattt gcagtgcatt 215640
agaatcttgc cgtcacttgc acacgtcacc aagttacttt agtgagagtt cagcctagct 215700
atggctctgc tgtgctaaca gttgcttttc aatattttgt ttgaggcttt ggaataattc 215760
aaaggcctac acttttttt ttctaatttg tttccttgga gttttacgca tggctacttc 215820
agaaaacgtc agttttatgt cattaatgtc atcatcttct ctggattctc agaattcaaa 215880
attcacagga gcatggcagc cttacattca gtctattctt ttcataaaaa aggaagtaaa 215940
ctgcaacagt tcgcctacgc tatggagact ggagtggtcc cacctctgta attctrtcts 216000
tgtctgcccc acagctgtgc cgaagygagt gccacttgtc tgcagggccg taccgcggaa 216060
ccctctttgc cgaccagcca gygatgtttg tctcgcctgc cagcagcccc ccagtggcca 216120
agctctgtga actagtccac ctgtgcggag gccgggtcag ccaagtcccc cgccaggcca 216180
gcatcgtcat cgggccctac agcggaaaga agaaagcmac agtcaagtat ctgtctgaga 216240
aatgggtctt aggtaagaat ccaggcacac agacgctgtg gtgtggtcca gatctgtgga 216300
caggtttcca gggagggcgg cktcaggctc acacccccct ccacgcagct ggggcacctg 216360
ggttgatgtc tcagcctcca gcatctgccc tggcagcgtc gtgtggtcac cctcggcatt 216420
cccgctcctt gctgttagca gacgtacagt tcacgaggaa atgggaactc tgactggact 216480
tccccacttg acttccctgg ctcgtgtgaa aaatccaggc tacccaaagc caccccrggc 216540
caccctgtg ggcacagact ctccgggcac ccctcttaga ccctccctcc ccagtgcctc 216600
cttgtcctgc ttcaggagtc cctggcagcg cccggcactg gggcccaagc ccccgtccct 216660
gtcatctcct ctcccaggta catctcatga tcactccgtc tgctcatgtg ctcaaagggt 216720
gttaaaagac gtcaaacgac tccatctttt atttgacaaa gtgagcacag tgtgaccgta 216780
atgtcccact ctggcgttca tggagctgcg ccaggcgccg tgtgcgattc tggggaggaa 216840
gaggtggtag gagctgagct gagatcggag gaggctggaa ccccacgccg tgctaacaca 216900
cgggctccag gagacttgca ggtgatcccc ggagaagagg gttaaggaag agtgtgaagc 216960
aaggacggcc tggggaatgc ggaggaagca ggggcagcgt ctgtgctaga aattacctgc 217020
cctgtggtgg agtcatatgt ggcgggacaa gcctagggct ccactgtggg gaaatcccac 217080
accctcctcc atggggttgt gataaacatg ttagtttgct tgggctgcca tcgcaaaata 217140
ctacaggctg ggtggcttca aacaacacgc attgtctctc agttctggag gctggaagtc 217200
taagatgggg tatcggcagc gttggtttcc cctgaggcct ctctcctggg cttgcagaca 217260
gctgccttct tcctgtgacc tcacgtggcc tttcctccat gcacacacat ccctggtatc 217320
tctgtgtgtg tccaaatgtt ctcttctcta aggataccag tcagattgga ttagggctca 217380
cccaatggca tacttttatt tgcttttatt tatttttttg aaacagtgtc tcgctctgtc 217440
acccaggatg gagtgcagta gcatgatcac agcttactgc agcctcagcc tctctggctg 217500
aagtgattct cctgcctcag cctcccaagt agctggaact acaggtgcac accacgatgc 217560
ccagcttttc tttctttttt ttttttttt tttgtagaga tggggtctcc ctatgttgcc 217620
caggatagtc tcaaactccg gggctcaagc gatcctcctg ctttggcttc ccaaagtgct 217680
gggattacag gtgtgagcca ctgcacccag ccccagtggc atcattttaa cttgtctttt 217740
tcaaggcccc atctccaaat acagtctcat cctgagttac tgagggttaa gacatcgaca 217800
tacgaatttt gggcagacac aattcagccc ataacaatga atcactctag tttcagcccc 217860
tggggccaag atccttaccc gactttagag gtacatcccc tctctctctc tcaatctctc 217920
tctctctctc ccgttctctc attctttttc tctctctttg cttccatctc cttccatgtt 217980
tcctattcag tctcctttct tagtactttt gcatgtctct aaatcctaaa cttctggctt 218040
ttctcatccg ctgctcaaca ttatccctta atagacaagt agatactgtg tttgttcaag 218100
ttacattcgt atctaactac ggacatttta caagtatctt ttacatgact gatggtcatc 218160
ctttcatata ttttagaagt gtggcaatca aaagtaattt tttactctgg tgcagagtaa 218220
ttcatctttt gcctggaaac caacttccaa aaaaaaaaaa actatgattt tagtcacagt 218280
ccaaaagcta agaggctgtt tactcttttc taaatgccaa gaatataacc ttcaaaacat 218340
cctatgttct gaaacagagg ttgttgtttt gtttttctgg agaagtgtat tatcaaaatg 218400
ccacggactg cagaacagaa ctgggcctga aagcatgtct gggccagctg acggaactgt 218460
gcacacgatt gatatccaca gtgcatatca acaggcagtc tttttggagt ttgcaaagcg 218520
tgtgccgtgc agtgcccgag cctgcctctg cactcgtgtt tccaggttgg gtggctctga 218580
cagcccctte ctgtgggtcc tgcgtccttg tgtggagtca cgcttgctcg gcagctgctc 218640
acttcctccg gttgttttgc cgctcggctc tcccgcccgt gggttttcag gaggcgaatg 218700
tctacctgct taatcctgag gcttcgatcc cgcaaagccc ttcagagttc tctgacttcc 218760
aggccctggc cacaggcccc agcctctttt tctttcctcc tgtaacttgt gtcctgtttc 218820
tgatttctca ccaattatgc catctgcctg tgcccttggt aacatctggg tattgtgtgt 218880
gctgcagacc tcacccatgt gagacaggtc ccctcactcg ccggccacca gaccccagtg 218940
tagtgggcgt ctccagcgta gtgggcgtct ccagtgtagt gggcatctcc agtgtagtag 219000
acctctccag tgtaccaggc ctctccagcc cacactctct gagatgtaag atcacgtagt 219060
tctcaagtat ttattggctt gtatttttct ctttgtgaag tgaattccaa tctagtagct 219120
gcagctatgt acgaataaag aagggtttat ttttctgtcc gtacatactt ctggcttttc 219180
tcaccctctg ctaaacatta tcctttaata gacaagtaga ttttttttgta tttttctctt 219240
tgtgaattga attccaatct ggtagctgcc gctatgtaca aataaaggaa ggtttatttt 219300

tctgtccata catacacacg taaacctaca gaacacacag tccagggcat tgcgtttcct 219360
gcctcatcca ggtccaggct atttgcttat tctctaacca gaaacaaatc atatactttt 219420
ttttttttt ttctgagatg gagtctcgct gtgtcaccag gctggagtgt gcagtgatga 219480
gatctcagct cactgcaacc ttcacctcct gggttcaagt gattcttctg cctcagcctt 219540
cccagtagct ggaattacag gccccgccac catgcccagg taatttttgt atttttagta 219600
gagatgaggt ttcaccatgt tggccaggct ggtctcaaac ccccaacctc aagtgatcct 219660
cctgcctcgg cctcccaaag tgctgggatt acaggcgtga gccaccgtgc ctggccgaaa 219720
tcacctattt tctgtggaat gcatttactt catgtataaa acagagtcat agcctccacc 219780
ttgcttaccc cacatgctgg ttaaaggagg aaacacagag agcgcaaatg ccctgtggca 219840
ggcgtaggct tcttaagtgt ggcagattga cggtatccat ggatgtgtcc tcatcatccc 219900
tgccccttcg acaaagcaca ttgtgtcttt tggagacttt ttttcctccc gttcatttcc 219960
attataacaa atgcttctct ggacaatgtt tcattctcaa aatatcgcaa tattgaaaaa 220020
ctaggaatat atcaaaccat tttaaagcac caaatcgaaa aagaagttat tttgtttaaa 220080
taaattatga aaagacaata ctcaaaaaaa aatcaattaa atttattcaa actggaatat 220140
caactgcttt gtaaggtagg gtccctgagc gtcttagagt aatttgagcc gggcgtggtg 220200
gcccatgcct gttgtcttag ctacgtggga gcttggcttg agcccataag ttcaaggctg 220260
cggtgagcaa cgatcccacc actgtactcc agcctaggca acagagcaag accccatctc 220320
taaaaagaaa aaaaaaagaa tcatttttca gtgcctttat attgtttctg tatcttaaca 220380
gtcttgtttt gcagatgtcg taaactcaca gggggtggag aaccaggagt tttttagcca 220440
ctaggaacct ctctgagaag tttcttttct tttcctttct ttattattat tagtattttg 220500
tggccagagg agggaaagga aggtgggtac tgaaacgaca gctcttcccc tgggactgca 220560
gcatccgagc accacagtcc acccgccagc ctttgttcct gcacagtctg cctctcaaga 220620
ccaacaactc catatctatg acgataaaaa ttgttagtga ttatttttact tgtaagaatt 220680
tctttcgacc tcagctctga ggtgaccctc agctcgcccg ccaccccagc tgccccacct 220740
tgctggcata gaacagggag tggaggtgtg aagtcactca acagggctca gtatacaaaa 220800
tgtaagccac gcctcactca cttgctccct ggagaatttc atctgcgccg cgttgcctaa 220860
taacggggtt atcggaaagg gcatgattac gttccctctt cattccctgg agtctttttt 220920
ccctgaaact gtattgtact tgggccaaga ttcttgatga atcattcaac cagaaggaga 220980
aatggggttg ttgtttggtt tttttgtttt gttttttttt ttttttgcg ttttgagaga 221040
gcacacttgt gggtggttga acatggataa aaataaacgg gaaaacaaaa atcaaattcc 221100
cggccctagg aaataaaatg ttacctttac ctgatattga taatacatat tatatttgaa 221160
agcatttgct aatggttgca ttttcccccc aacactccca tgacatataa ttcccatttt 221220
ataagtcacg aaacgaagac cctggggtct gaaggaactt ggctggggtg aggatcacaa 221280
gcccttgggt ggagctctga gccctggcgc ggtcctcaag ggtctgcgac atttgtgctg 221340
tggtcagctc tgtgcactct tccctccctg ctgctgttat cacgaaaggc tggcttggcc 221400
tttctcatag gcgtatttcc actctcaggc gccctttat tgtctgggct ccattcaagt 221460
gataagacat acatttatgc tattgtggga acataatgta atattctcaa cagcattgcc 221520
aaacaaaaaa aaagtttagc ctctgcctga ttttcttata acttataaag aaaatttggt 221580
ttgaacatgt cccatgtcga tgttttcagg aaaaagatcc gatagcatgc aggccttctc 221640
atgctggcmt ggctcattca tcgtttcccc taatgactga ctgaccagaa aaatgcacga 221700
cgctcccatg gggccactcg ggaggcctca ggcttcgggc ttcctgattc agtagatatg 221760
tgaggcttga tcagtcaccg cagtccacat ctccattgcc tcgataagga accagtcgca 221820
gagaggggag gccatctgca gaagctgtgg agagtggcag agaggaragt gaggacgggg 221880
actgcccccc tccagcccct ctcctccaag gacggcctca ttttatcccc acccaggttt 221940
ccacacccag gagctcagca accgctcaga aaatgtttgt agaattcaaa gacataattc 222000
agacaatatg aagaattatt tttcctttga gttgttctta aaacagacga aatctaccag 222060
catataaatg aatgagaact aaaactggtg ggatttggta atgtcgacat ctgagatgtt 222120
taggctttta aatatatatc tcagccaggt gcggtggccc atgcctataa tcccagcact 222180
ttaggaggcc gaggcgggtg ggtcgtttga gcccagcagc tcgagtccag cctgggcaac 222240
atggtagaat ctcgtctgta caaaaaagta caataattag cgggcatggt ggtgcaagcc 222300
tatagttgca gctacatgag aggctaaggt gggaggatca cctgagctca gggaggtcag 222360
tgctgcagtg agctgtgatc atgccattgc actccagcct gtgcgacaga gtgagaacct 222420
gtctaaaaat atatatgtgt ttatatatat atatttatat aaacattagt gggttttaaa 222480
aaaaattaac taactgctag ctcctaaaac agtattttgc cattagcttt ggaaaggttt 222540
gctcagaaaa tgaatttcta agcactccct tcattgcatt tattggtcaa actaatggtc 222600
ctggatggtt atctttgaaa cttcctaacc tgttgggtcc ccgtcgttaa acttatgcca 222660
acagaactaa actcactgga tgtgaattgc atcagagatg taaacattta aaagcgtatt 222720
aaggctgggc gcagtggctc actcccgtca tcccagcact ttgggaggcc gaagcgggcg 222780
gatcatgagg tcaggagatc gagaccatcc tggctaacac agtgaaaccc cgtctctatt 222840
aaaaatacag aaaaattagc cggtcgtggt ggcaggtgcc tgtagtccca gctactcagg 222900
aggctgaggc aggagaatgc atgaacccgg gaggcagagc ttgcagtgag ccgagatcac 222960
gccactgcac tccagcctgg gcaacagagt aagactctgt ctcaaaaaaa aaaaaaaaaa 223020
aaaaaaacat taaaagcaga ccaagaaaat cctagaatac aggagtcagc tgtctattca 223080

attcagaata agaaatattg tagacaaggc aacattttat gtgtattaga aatgtggtgg 223140
ttggtttgag aagtgaaacc agccatgtat atgctgctcc aagcattttg gttgtggcag 223200
gaaactttga agactatttt gctgtacaaa ttcacaaagc cccctgcaaa cactcccgtg 223260
cttggggtga atgcccaagt gtgtcacagc tgccttgcag ctctgaggat cagaaaggtt 223320
aatggacata aaagaaactt caaagctcaa cctcctaatg ggaagctgcc cttggtttta 223380
ggctgtcttt gcttactgac cgacttaatt catgctttgg gttatgactg taggagagat 223440
tttcctgtgt ctttggagta tgctgaactt gtgtttcttt ttgttgttgc atattagaca 223500
gtcagtgttg aaactaaagt gacctaaagt gacagagctc atgttatggg ctgaattttg 223560
tctccccaga attcataggt tgaagccttc ccagtcctta gaacatgatt gtatctggag 223620
ctagggcctt taaagacata aataaggtaa catgaggtca taagggcaag gccctaatcc 223680
aatatgactg gtgtccttat acgaagagga agaggccagg cgtggtggct tacgcctata 223740
atcccagcac tttgggaggc cagggccggc agatcacttg aggtcaggag tttgggacca 223800
gtgtgtccaa catggtgaaa ccccgtctct actaaaaatg caaaattagc tgggcatggt 223860
tgtgggcacc tgcaatccca gctacttggg aggctgaggc aggagaatcc cttgaacaca 223920
agaggcggag gctgcagtta gtcgtgatcc caccactgca ctccaacctg tgcaacagag 223980
caaaacccca tctcaaaaaa ataaaaataa aataaaggaa gacaaagaaa caccaaagat 224040
atttttgcac agagaagagt ccaagtgagg actcagggag aaggtggcca tctgcaaccc 224100
gagcagtctc ccaggaagcc tcaggagaaa ctaacccctg tgacaccttg gtcttggact 224160
tcctgccctc cagaactgtg aaaaaatata tgtctgctgt ttaagccacc caccctgtgg 224220
cattttgtta tggtagcctg agcaaactag ttcagcccaa aatgaattct gatatcacct 224280
gcagaaatct gcttttagac agcaggaaac tgagggcctc tgagtttcta ggccagagtc 224340
atgcagtgaa ttactgaaag acccagaacc ccagtcctgg cccctgattt tcagtttaga 224400
atcttccttg gtaagaagca ggatcttagg ctgggcccag caagtggaaa actcttttt 224460
atttacacag ccactgactg ttgtggtctc agactgtacc acagaacctg gtgttccaca 224520
aacttcccca gtttggagca agagaaaaaa gtagttggat gaaatgatct cattttattt 224580
tttagtcaat ttttcttaaa tgttggtgct tgaaaacaaa tggatggcag taaagtaatc 224640
ctgaagaaca caggaggaaa gaaataaaag aggcaatacc aaatgttagc aaaatggcag 224700
caaggcaaat aagaggctca gcaatagcaa aaaactgagt tctttggctg ggaaaaactt 224760
ataaatatta aaaatcctga caatgttgaa aaagaaaggc agagataggg ttccaggaga 224820
aatactaaga atgaaattgg agctgtcact gcagttatcg taaggatatt ttaaaatcat 224880
aagagagcat gatgaacaat ttaataccaa taaatttgaa aacaggtaag atggatgatt 224940
tttagaaaaa tgttaccaaa attgattcaa gaaatagaaa atctaaacaa gctcaagcgt 225000
taaaaaaatt aaataggtaa aatatgtaca tcaactgggc acagtggctc acgcctgtaa 225060
tcccaacact ttgggaggct gaagtggaca gatcacttga ggtcaggaac tagagaccag 225120
cctgaccaac acggtgaaac cctgtcttta ctaaaaatac aaaatgagcc aggcatgatg 225180
gggcatgcct gtgatcccag ctacttggga ggctgaggca ggagaatcgc ttgaacctgg 225240
gaggtggagg ttgcagtgag ccgagactgt gccattgcac tccagcctgg gcaactagag 225300
caaaactctg tcctaaaaaa aaaaaacaaa aaaaaaacaa ttatatatca acaaaaaaaa 225360
gaaaatttta aaaagtaaca atttgaaaaa gtcaaatagg caatcaaaag tattcctttc 225420
accagccact aaaaaggcac ctgtacatgg gaatggtagc aaaatgacag aagaggaaac 225480
tctaacctct catccaacac agaaaccgct aaaaccaggc agaagctgtc tgcagagatg 225540
ttgcaggtgc tctaaaaggt gctctaaaca accaccaaat gcatacagca accaggcaaa 225600
tgcctgatag aggaaagcca tcttcaagcc cgcaggaaag ttttstggca catggtggca 225660
acccagttcc cagttcccag ttcccttcct caagctgcag ggagcagacc agacatgatt 225720
tgttctagtc tagctgattc atacctgaag gattgatcct catctccatc tcacataaca 225780
tgcaaggtgg gcaagaaaaa gaggtgggca cagctcatga aagccacaga gaggcaatta 225840
aggtaaaaat agataaattg cactatatac aaattaaaga cttcagtgca tcaaaggata 225900
cagtcaacag agtgaaaagc aatctatgga ataggagaaa atatttgcaa ataacgggtt 225960
aatcttcaca atatataaag aactcctgca actcaacaac aaaaaaaaac cccagtttca 226020
aactgagcaa agaacttgaa taaacatttc ttcaaaaaag atgatataaa tgtccaatag 226080
gcaaatgaaa agatgcttaa cattactaat ccttaggaag atgcaaatca aaaccacaat 226140
gagatagcac ctcagcacct cacacccatt atgattgcta ctataaaaaa aaaaaaaaac 226200
ccagaaaata acaagtgtta gtaaggatgt ggaaaattgg aaccttgtgt ctgcctcatg 226260
taatgttggg aatgtaagat attgtagcca cgatagaaaa cagtgtggca gttcatcaaa 226320
aaatgaaaag tagaattact gtatgatcca acaattcctc ttctgggtat atgccaaaaa 226380
aattgaaagc aggatctcaa aagaataatt gtacatccac atttatagca gcattgttca 226440
caatagccaa aaggcagaag cccaagtgtt catcagtgga tgcataagaa acaaaatgtg 226500
gtctatccat acagtggaat attattcacc cttaaaaagg aaggagattc tgatacatgt 226560
aacactgtgg atgaactttg aaaacatcat gttaagtgaa ataagccaga aaccaaagga 226620
caaatatcat acgactacac ttataagagg aacttagaat agacaaagtc acagagacaa 226680
actatagttg aattaccaag ggtggagtag gcaggaaggg agtggagaat tattgtttaa 226740
tggctacaga gactcagttt tggataatga gaacattcta gaaattaata gtagtgatgg 226800
ctgcacagca ttgcgaatgt acttcatgcc actgaagtgg acacttaaaa atagctaata 226860

tggtaaattt tatgttatgt ctatcaaact tttaaaggca ccctccacag atagttttag 226920
tagtaagttt taccaaacat tataaagttt tacaggaaaa aaaaagaaat ctattcacct 226980
cattttacaa ggctacattg atcttgacct aatactggtt taaaaaactc atttgtaaac 227040
aagtacataa aaatctgagg ctgagcgcag tgactcatgc ctgtaatccc aacactttgg 227100
aaggccgagg ggggcggatc acaaggtcag gagatcgaga ccatcctggc taacacagtg 227160
aaaccccatc tctactaaaa atacaaaaaa ttagccgggc gtggtggcat gtgcctgtag 227220
tcccagctac tcgggaggct gaggcaggag aatcacttaa acctgggaga aagaggttgc 227280
agtgagccaa gagtgcgcca ttgcactcca gcctaggcaa cagagtgaga ctctgtctaa 227340
gaagaagaaa agaaaaaaaa actcagaaat aagatatttc atcaagtcaa atttggtagt 227400
gtgtttttaa aacacacaca cacataacca agtgtggttt aacctaagaa tgaaaggata 227460
aatgaatagc attaagtctt ctttttttcta atccattaat tttcttagta gtgttaaaaa 227520
gcagtaggga agattcaatg ccgagtaatg atttaaaaaa aaaaaaactc ttcagaaacc 227580
aggaatagat aactttctta actatggagg ttatctataa aaaacgtaca acaaatattg 227640
aatggtgaaa accttagttt aaggcttaaa tcaggtacaa gacacacatg aatgctatta 227700
ctcttcaaca gtgttctatg attcctagtc aagggaataa aataaaaaaa attacaagaa 227760
ttatacagga agggacacat tttgtttgca tgtcatacag ttgtctacat agaaacatca 227820
aagagagtca ataaactgtt acaactcatt cagcaaaatt cctctttgta agatccactc 227880
actgaaatct ttagcatttg tatacccaat gataaacaat tataaaatgt aacagaaaac 227940
atagtaaata atagtggatt caaggctagc catgtaatac agattgaaca ttcctaattt 228000
taatctgaaa tgctccgata tcttaaactt tttgagtgcc aacctgtcaa cacaagtgga 228060
aaattccaca cctgacctca tgtgacaggg catagtcaaa gcacaggtgc acgacacagt 228120
tgatttagcg tccccaaggg aaaaaaaaga cccacccagc ccccttcaac tatagtataa 228180
cttttccacg cacacccaaa ttcccccaca caagcacgcc cacaatgtgt aataaaatgg 228240
cacgtgtgca ggctggacgc acccaacgca gattccccac gatacctcac gtggggccga 228300
gaactccatg cattactcac tgtggttttt tgcttattct ctgcagtgtc atgtaaaaat 228360
attactgaaa atgtcgaaaa ggcctgcaga tccccctatg tgtaacagtg atcagaaaaa 228420
gaggaataat ttatgtttat caatagcaca aacagtcaac ttgttggagg aactgaacag 228480
cagtataagt gtgaagcgtc ttacagaaga gtatggtgtt gggatgacca ccatacatga 228540
cctgaagaaa cagaaggata cgcttttgaa gttctatgct gaatgtgatg agcagaagtt 228600
aatgaaaaat agaaaaactc tacgtaaagc taaaaatgaa gatgtgaata gtgtattgaa 228660
aaactagatc tgaaggcatc acactgaacc cgtgccactc agtggtaggc tgatcatgaa 228720
acaagcgaag atctatcctg atgaactgaa aattgaaggg aactgtgaat attcaacagg 228780
ctggttgcag aaatttaaga aatgacatgg aattcaagtt ttaaagcatc tgcagatcac 228840
aaggcagcgt cgaaactcat tgacgagttt gccaagatta tcgctaatga aaatctgatg 228900
ccagaacaag tctgtattgc tgatgagaca tgaccatttg ggtgctactg ccccagaaag 228960
atgctgacta cagctgacgg gacagcccct acaggaatta aggatgccaa ggacagaatg 229020
actgcagtgc tgtgcaaatg cagcaggcac gcataagtgt aaacctgctc tcatgggcaa 229080
aagcttttgt ccgtgctgtt ttcaaagagt aaatttctta ccagtccatt attatgctaa 229140
caaaaaggca tagatcacca gggacatctt ttctgatcgg ttttacaaac acttcgtaca 229200
ggcctcttgt gctcgctgca gaaaagttgg accggatgat gacagcaaga ttttcttatg 229260
ccttgactac tgttctgctc atcctccagc tgaaattctc atcaaagata atattgatgc 229320
tgtgtacttt cccccaaacg tgacttcatt agttgagcct gtaaccaggg tatctttaga 229380
tcaatgraaa gtaaatwtaa aaacactgtc ttgaattgca cgctcgcagc agtgaacgga 229440
ggtgtaggtg tagaagattt tcaggagctg agcatgaagg atgccataca tgctgttgcc 229500
aacacttgca acacagtgac taaagacaca gatgtgcgtg cctggcgtga cctctggcct 229560
acgactgtgt tcagtgatga tgatgaacca ggtggtggtt tagaagaatt cagcttgtca 229620
agtgagaaga aaaggatgtc tgacctccaa aaaatatacc ttcagagttc atcagtcagc 229680
gggaagaagt acacattaat gtcattttta acattgataa tgaggctccg gttgttcatt 229740
tcattgactg ttggggaaat agccagaatg gttctgaatc aaggtgatcg tgatgatacg 229800
accatgaaga tgacgttaac actgcagaaa aagcacccgt ggacagcgtg gagctcaggt 229860
gtgatgggtt aactgaggcc cagagcagcg tgcattcaca acagaacaag caatcatgtc 229920
agcttataaa atcaaagaaa gaatcctaag acaaaaaaga aagaaaaaaa attagccggg 229980
catggtgaca cgtgactata gtcccagctg tgtgggaggc tgaggtaaga gtcttgcctg 230040
agcccaggag ttagaggctg cagtgagccg tgatcatgcc actgcacacc agcctgggaa 230100
acagcgaggc cctgtctcaa aaaaacccaa aaaactaagt aaatattttg tacatgaaac 230160
aaactttgtg tacactgaac caacagaaag cagctgtcgg ttctgagacc attgttagtg 230220
gtgcagatac cattaaaaag ccccccagca gaatgcctcc tcgtccccag aggacccact 230280
tcctgggcct gtaactgctt cttatgttcc ttctcaccta aaatgtaaaa tgccgtgtcc 230340
cgtaagcttt gaatcaaagc acagcatggt tgggagagca gaggcctgct gttgtttgtt 230400
gttgctgctg ttgttcagca gctgattgcg gtctctgctg atgccactgg ctgcttagct 230460
cccctgagca cgtaagtctt cactgtgtta atggcatgtc ttatttttta ctgtgaagta 230520
cttatgtgtg aataagtgta aggaaatgac tgcttggtag tagcatataa attcagagtc 230580
acgggcaggc acggtggctc acgcctgtaa tcccagcact ttgggaggcc aaggcgggca 230640

gatcgcttga ggccaggagt tcaacaccag cctggccaac atggcaaaac cccatctcta 230700
ctaaaaatta caaaaattag ccgggcgtga tggcacatgc atgtagtccc agctacttgg 230760
gaggctgagg caggggaatc gcttgagcct gggaggtaga gattgcagtg agccaagatt 230820
tcaccactgc actccagcct gggtgacaga gagactgtct caaaaaaaaa aaaaaaaaag 230880
tcacagtcag gaatgagggt gatgccacac aaccactgat tgtccacatg ggggtgaggg 230940
ctgagatagt gatacctctg ctttctgatg gttccatgta cacagacttt gtttcatgca 231000
caaaatttgt ttgtttattt tttgaaacag agtttggctc tgttgcccag gctggtgtac 231060
agtgctgcga tcatagctca ttgcagcctt taactcctgg cctcaagcga tcttcccacc 231120
tcagcctccg ttgtagctgg gactacagtc atgctgtcgc acctggcaat cacaccagtc 231180
tatgcacaga actatttaaa atactgtata aaattacctc taggctatgt gtataagatg 231240
cagatgaaac ataaatgaat tttggtttta gactctggtc ctatcttcaa gatctctcat 231300
tgtccattcc aaaaatgcca cccaccaccc cccaaaaaaa atctggaatt caaaacattt 231360
ctggtctcca gcattttgga taagggacac accacctgta atatcctttt acacatttcc 231420
tggatgggaa acagaagttg gtgtggtagg agtcacacat aaacggcaga ctttcttgtc 231480
tgtgacacat tcttaggatg tcctagagaa gtatcagcga tgtgaatgtc tccagtcaaa 231540
tatcagagca gaaagaatat gttgagaact gctgtattat tagactgggc tactttcttc 231600
aaacaacaca tggtatcagg tcattcattc atttacccag tagatatttc ctacacactt 231660
gtcatatgcc gagcatatcc taggcactgc aggtacagca actgacagga atatacagcc 231720
tttgcccttg tgggacttaa catttaagag agaagacagg cagcaaacaa tttctttaaa 231780
aatccttctg gtggtaaatg caatgaagaa aacagggtga gtatagagag gaggagtgag 231840
gtaggcccct tgcacgtgag tggcatttga gctgaggccc agatgatgaa gagaaggatg 231900
gactcttgta ggtctattgg actggccctt ccaggaatgg taagggctga gaggtcagga 231960
gaagcggtaa gtttagcgtg gctgaaatga agggagagaa gacaaagcaa taggaaatga 232020
agctggagaa gcaggcagct tcagacagga ccattccaga ccactgacac cttaacagac 232080
aacagcaaga agtttgggtt ctgttctaag gataaatgga agtcacagaa cgattttaag 232140
tgggaggatt aggctgcggt atatgtttgt ttactctgtt tgtgtttatt tttgttttaa 232200
tggatacaga gtctcccaat gttgcccagg ctggttttga actcctgggc tcaacggatc 232260
ctcctaactc ggcctctcaa agtgctgaca catgtttttt taatggaagc agagaaagca 232320
gtctcggacc tttgcagtgg ttcaggtgat tagtgatggg ggttaggacc agggacgtat 232380
cgatggaggt gttgtgaagt tgtcatattt taaatataca tttcagagcc aggtgcactc 232440
gctgatacat tggatgtggc atattagaga aagaagactc gaaggtggca cctagtcttg 232500
tgttctgagc ttccagaatg aggcatctag aagccaggac ccgggagaag cacggaagga 232560
gcagtggttt attcagtctg cagaagcagt gcctacagga ctgctgtgtg aaagaggaca 232620
catgtgatat gagcagatga aaatcacaca gcaggcagct ctgggctcat tatgagaaac 232680
gactctagga atatttgtaa cctgctgggc tctactgcta agggctgcct taagccatga 232740
agccgcagag gctgggtgac caccgtccca cagtgaggga gctgggcaat tccttaccag 232800
agtggagatg tggctagatc tcctagccct aacatgctta cttattttga taagcaaaga 232860
tgaagctcac atgggtcccg tgtgctcttg aacttctgta cattgtacca ttaaccacac 232920
ttggatgctg gcaatcgcag ttttagttaa ataaagtgac ttgcccacca tactataaaa 232980
aattaatttt ggtagcatgt tgattctgta tcctaaccat aagaccacac agagccatgg 233040
ctagtaaact ttagcttgtg cgtaaatgcc tgccaagacc tgctaaatac tgttgcttac 233100
atttaaaaaa aaaaaaaaaa ttttttttt aatttaaatt tcacggagct gctcaagggc 233160
agttcagctt cctattcatc tctgtctcca ccggccagga ctggcattac tctaacatct 233220
gtctacggcc acattttatg ggatgtttga ggattattcc tatgaagtga cattggaatt 233280
tggggatgtg gctatgttca gatgccaaat aaacttggat agaaatcatt tttcctgtgt 233340
gtgtttacag ttaggaacgt ggggctgtga ggggctccct ggacatgacc ctggagctgt 233400
cggcccttgt tcagtggtca gatgcgcttc agacctccca gagtgctgcc cgcacactca 233460
gtcacagccc catgcgcacc tcaacgccac tgctcagaag tccagtgtaa ttcctcaggc 233520
agcatgtcct agagcaggcc atgagaggtg taaggtacag actttgttgt gaggttacat 233580
gtaggcttct gttccatctt gtctctgttt aaagatcgat acttctggca gcctttatcc 233640
ccaccacgat aaatacgtgg atggaaggat acatgcgtgg aagggtggat gggtggatgg 233700
ttggatggat gggtagacgg gtgcatgggt agatgggtag atgggtggat ggagtgatat 233760
ttgatttcat agtcaaagaa ctcaaacagt agacaagtac acagggtcct ccagtcttac 233820
aacccttcct taactacaat aaagatagaa gtgtatcttc tagatttctt ttaaaaacat 233880
atttatgaat gtaaacatat tatggtcagg tccagtgact cacatgtata atctaacact 233940
ttgggaagcc aaggtgagtg gactgcttga tgccgggagt ttgagaccag ccttggcaac 234000
atagaaagac cgtgtcccta caaaaaaaat tttaaaagta gcctggtgtc atggcacatg 234060
cctgtagtcc tagctactca ggacgtcaag gtgggaggat cactcgagga caggaattcc 234120
aggctgcagt aagccatgat cataccactg cactccagtc tgggcaatgg atcaagatcc 234180
tgtctcttta aaaaaaaaaa aaaacatatt tacatagaaa taaatgtata taaacacaga 234240
tattgtttag ggatttgttt ttatatatat tggagaatga catgcttttt caggagcttt 234300
tatttaaccc tatgcctaga agatccttcc agcttaacac atatacagct acttcattct 234360
ttttaaccat tgggaggtac tgtaattaat ttatgtgctt tctgttattt tcattgtttt 234420

gctattgtat ttacttattt attttagaaa caagatgtca ctatgttgcc caggctggcc 234480
tcaaactctt gggctcaagc agtcctccca ccttagcctc ccaagtaggc gggactacag 234540
gcatgaaccc tgcaatacgg ctggcttctg ctattttaaa ctcgtgtgtg tgtgtgtgtg 234600
tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg cgcgcgcgtg tgtgtgtgtg tgcgtgtgtg 234660
tgtgtgtttt ctaactgaac aatctgaatt caattttaag agattttctt gagctggaat 234720
tattctagtc cgagcccagg ctcatgaaga tttctgtaaa atacattcca agcagtgaaa 234780
ttactgtgcc ctaggatatg tgtacttaaa ttctgataca caaggctgca gcaatttaca 234840
ctattactaa cggtacataa agtcctattt cctatgtcct ataaattccc atgtccagta 234900
ctggacataa cccatatttt caatattggg tgatccgatt agttaaaaaa atagatctca 234960
ttaatttcta attgcctgat tactaaatta tgaatgagtc tgaatatctt agataggaga 235020
tttatcattc gtgaattacc tgtcctgatc ccttaactgt tttgaaattg ggttatttat 235080
atttttcaca tggttttaca gcaatgttta cataatatgg acattaaact tttgttgtgt 235140
tataaaactc tgtctcttta gctgtgctta tggtgtctta agtattacca agtttttaat 235200
ttttaactat tatttttac aaaattaaac acctcttttc ctccatggca cctacccttg 235260
tggttttgct tagaaaggcc ttcctcaccc tctgagcttt aaaaataatc tcatattctc 235320
ctatttatag ttttaaaaaa tatttagacc tttaatgcat gtgcatttca cttactgtat 235380
aatgtgaggg gaccatgttg tttttaataa ctaatttatt gacactgacc tatattgccc 235440
cctgtgagtc atctcttaca ttcccacatg gtatgggtgt gtttctggtt attctcgtcc 235500
attgatctgt ttgtctattc tgtgctgacc tctattttac tgctataatt gtacagactg 235560
ttttgatatc tggtatgtca aatttttttct catcatttct ctttttaaaa atcatcttcc 235620
tatgcatttt tttctttcct ataaacttta gaataaacat gtcgttttct ttttgaaaag 235680
tttgaaattt ttggattaca ttgaatttct agatgaattt ggaaagagca tcatttttc 235740
tgcatttttt tatgattttt caaaactgac acctagtcag aaaactaagt gtaaaaattg 235800
aatccataga gtttttacaa cctggaagaa aatacaaatg tggctgaatg actttaaacc 235860
ctgagtatcg gaaaaggctt ccacctacct atgactcaaa agccagatgc aataagacaa 235920
agtgttgata taatttgaat acataagaaa ttgaaactta tacatggcaa aagtttgcat 235980
aagaaaagtc aagccaggtg tggtgggtta tgtctataat cccagcattt tggaagactg 236040
aggcacagga agattgcttg agcccaggag ttcgagatca gcctgggcaa caaagtgaga 236100
cattggctct acaaaaaatc aaaacattaa ctgggtgtgg tggtgcatac ctgtagtccc 236160
ggctacctgg gaagctgagt ctggaggatc acctgagtcc aggagactga ggctgcagtg 236220
agtcatgttt gcaccaatgc agtctaacct gcgtgactga gcaagaccct atctcaaaaa 236280
aagaaaaaat atgtaaatca taataatacc tgcttcactg ttgtggagag aattaagtag 236340
tatgcctagt actaataata ttgttataat tatatacaat gtttttaact atatcatttc 236400
ttatatatat aagctatcac aaatgttagt gttcctccct tctgaaattc atctgagggt 236460
ccctcactga cccaggcctc ctgggtagaa gcacatttgt attgagaaga caacagttaa 236520
attctgggac actatcttga gctataacta agataagtca ttttttttctt ccatttctaa 236580
aaatatttgt agattaaacc catttttttc ttttttgtac cataccacca ggatagcttt 236640
ccaccttcca tcactcatct gtgtgacttc ttaagttcct tcaaatgtaa ctctgtaatt 236700
ataattatat attcacacaa tcattgtgat tctttaattg caattgattt aatctacctt 236760
atcatccaat cggtgctgac agtggatttc attccttttt ttttctaaca gtaggaatag 236820
aatgcagtgc gcttgccagg actgaggaaa gagggagggg ttgtttccgc cagctgccag 236880
gatcacctgt gctgaccctt cagcagcacc tgcagcgcta tcctgggcca ggcgcaactt 236940
gtgattttca taaaatagtc gagtttcaaa cggatgggac tttagagctt ctttaatttg 237000
agctatgaag aacagagttt tagaaagtat gcttattcac ttggaattcc ataaaaaata 237060
cctatgctgg gtagatagga tagcacggcc tacctctcac cactggtgtc ataattaaaa 237120
ctcatatatg tatttactta tactctgcct tatgccaaga gtactggaag tggtgagcta 237180
agattagaaa ttcttggctc ctatgtcaca gactggcaag cttcccaccc tgcccactga 237240
gtgtcctgac acaacgggaa cgtgccctgc atctaatggg acatgtggct accaagcact 237300
tgaactggcc agtgtgactg agaactgaat gtttcattgt attgaatttc gtttcacgtt 237360
aatttaaaaa ggtatgtgtg ctctatggac gtgggggggc ctatggacaa cacagctctt 237420
ggctatttgt ttttaaatat agtttcatgt atatacaaac aggttatcac tttcctatgt 237480
ggctggctat tatgaatgct aaactgcttt tcgctctctc tctagattcc atcacccagc 237540
acaaggtctg tgccyctgaa aactacctat tgtcacaatg acagtgacct cactggcctg 237600
tggtgactgc acacagctcg caaaactgtc tttggatgtt caaatgagaa acaaaactgt 237660
gaagagaagg aactggcgta tacaagatga cttctgatat catgtttgcc atgtgttgtg 237720
gttcttaaga actcataggt gactttctga tgactgaatg tctgtttcag agacgcttcg 237780
ggccttttta tttttatttt atttttttatt ttttgagacg gagtcctgcc ctgtttccca 237840
ggctggagtg caatggcaca atctcggctc actgcaacct ccacctccca ggttcaagcg 237900
attctgctgc ctcagcctcc tgagtagctg ggattacaga tgtgtgccac catgcctggc 237960
taatttttgt agttttagta gagacagggt ttcgccatgt tggccaggct ggtctcaaac 238020
gcctgagctc aggtgatctg tcaggcctct tctatagaat tccagtcttt gtgtcttagt 238080
catgatcata attgaaaggt cacagaacct ttgtcattag agcacagtac tgccaaataa 238140
agaatggaaa ttcaatgaca ttgttttatt actgagaaca actagagaac tctgcaagtt 238200

tcttggctta gactcgatct ttattaatac attatctatt aggtaggaaa gacatttgtc 238260
agctattaag gtgactttta tctagcggag attcctctct taaagtaatg aaaggagata 238320
ggtatggggg gtgttataca ggataattgg tgacatctga gtgtcttact tctgcaagcc 238380
tgctttatgg tgagcaaagc atcaccagca agtgatcaca atgtccactg gccgcttttt 238440
gcctgccgtc ctcgagatga aattggcagt tggggctgat tcacagaaac accgatttgt 238500
ggctgagcac ggtggctcac acctgtcatc ccagcccttt gggaggctga ggtggacaga 238560
tcacttgagg tcaggagttc gagaccagcc tgaccaacgc agcaaaaccc atctctacta 238620
aaaatacaaa aatcagctgg gtgtggtggc acacacctgt ggtcccagct cctcaggagt 238680
ctgaggcaga agaatcgctt gaacccaaga ggcagaggtt gcagtgagcc aaggttgcag 238740
tgaatcaaga ttgctccact gcactccagc ctgggcaaca gagtaactct ccttctcaaa 238800
taaataaata aataaataag aaacactgat gtgtctgtca ccttctaaag aaatgaaatg 238860
ctaggaagtc ctagccagag tgatcaggca agaataagcc ataaaaggca tccaaatagg 238920
aaaagaagtc aaactgtctc tcttcactgc cgatatgatt ctatacctag aaaaccctaa 238980
agactctgcc aaaaggctcc tggaaccgat aaatgactta agtaaagttt caggatagta 239040
aatccatgta caaaaatcag catttccaaa cacagtaaca ttcaagctga gcaccaaatc 239100
aagaacgcaa tcccatttcc aatagccacg gaatgaaata cctaggaaca cgtataacca 239160
aggaggcaaa ggatctctac aaggagaacc ataaacgaga tgctgagtcc cagcgaggtc 239220
ggaggtgcca ctgagccctc atcgtggtgc cgttcccgct ctgggttatt tatctgttgc 239280
tcatctcagc tgttgttcct acctcaaatt tcaagtccct caacaaatat aacagaacca 239340
cttctagaat gaacctttga gaagggaggt agcagtgcat tgtataggaa ttggcattct 239400
atagaaaacc acagaaactg gaaataatga agggttgtct cttggtttta aaataatgta 239460
tacacctaaa tcatcccctt atgatactca tcctctaaca gcaattgaac ttcaatacaa 239520
tgagtcattc ctgagttcac tcgcttcaca ttacatatgt ttctctataa ccacaagcat 239580
cctggcttgg tagtgctccc acagcaccaa aaatccctga ggaggctgac aaacattgtg 239640
ctgactcatg ctggagacaa gccacagaga acttccatcc cccaccacat cagccacgga 239700
gccagcccag cctctgccca cccaggcctc agtccccagt gttaagttct gatccctgat 239760
gctggcctgc cagtggccag tcaagattct ctttctgaaa gctagtattt tatgaggact 239820
gactgttgct agacattaca ctaagcacat tatatgttgt acttcatttt acccttttcaa 239880
caatcctatt agtagcttac tgtgggtctg caaagcctta ctcaaaacat atagggctag 239940
aggttctcag gattctgaat tttaaaaaaa atttgtaaag gcttatggct ctcaccactg 240000
ttattcaacg ttgcattaaa gtttctaccc agagaaggca ataaaaggaa attaaagcta 240060
tacagattgg aagtgaagaa ataaaagtct ttattctcaa gaatacaaga cactatgtat 240120
agaaattgta aggaatgcaa aaaaaaaaaa aaaaaaaagc cctacaagaa cttataacaa 240180
gtttagcaag attgcaatat acaatcttgc aatcttccta aagattatat acaaacctaa 240240
cagaattgta tttatatata ctgtcaataa gcaattcaaa atgaaattaa gaccacgatt 240300
ccatttaaaa ttgcatctaa aaataaacaa aataggaata gacttggcaa cagttgtaac 240360
atctgtatac tgaaacctgt aaaacattgc tgaaagaagt taaagacttc tttaaataga 240420
gacatataca aagttcatag attagaagat gcaatattgt taagatgata gtcctcaaat 240480
tgacgtatag attcaatgca atccattaaa atctcagatg gctttttata gaatttgaaa 240540
agctgatgct aaatctttta tgaaaatgca aagaacctct agtagacaaa acaatttttt 240600
taagagcaaa gttggaggat ttatagaacc tgattccaaa actgtcagta aaactacaat 240660
aattacaaag tatcagccag gtgccgtggc tcacatctgt aataccagct ctctgggagg 240720
ctgaggcggg tggatcactt gaagtcggga gtttaagacc agcctggcca acttggtgaa 240780
accttgtctc tactagaaat acaaaaaatt agccaggcat gatgg 240825

<210> 2
<211> 3809
<212> DNA
<213> Homo sapiens

<220>
<221> 5'UTR
<222> 1..57

<220>
<221> CDS
<222> 58..2565

<220>
<221> 3'UTR
<222> 2566..3809

<220>

<221> polyA_signal
<222> 3795..3800

<220>
<221> allele
<222> 285
<223> 5-392-222 : polymorphic base G or T

<220>
<221> allele
<222> 968
<223> 4-58-318 : polymorphic base G or T

<220>
<221> allele
<222> 997
<223> 4-58-289 : polymorphic base G or C

<220>
<221> allele
<222> 2102
<223> 5-398-203 : polymorphic base A or C

<220>
<221> allele
<222> 2283
<223> 5-400-175 : polymorphic base C or T

<220>
<221> allele
<222> 2339
<223> 5-400-231 : polymorphic base C or T

<220>
<221> allele
<222> 2475
<223> 5-400-367 : polymorphic base A or C

<220>
<221> allele
<222> 2539
<223> 5-402-144 : polymorphic base C or T

<220>
<221> variation
<222> 345
<223> polymorphic base A or G

<220>
<221> variation
<222> 615
<223> polymorphic base A or G

<220>
<221> variation
<222> 663
<223> polymorphic base T or C

<220>
<221> variation
<222> 666
<223> polymorphic base T or C

<220>
<221> variation
<222> 853
<223> polymorphic base T or C

<220>
<221> variation
<222> 989
<223> polymorphic base T or C

<220>
<221> variation
<222> 1309
<223> polymorphic base T or C

<220>
<221> variation
<222> 1472
<223> polymorphic base A or C

<220>
<221> variation
<222> 1839
<223> polymorphic base A or G

<220>
<221> variation
<222> 1913
<223> polymorphic base T or C

<220>
<221> variation
<222> 1998
<223> polymorphic base A or G

<220>
<221> variation
<222> 2319
<223> polymorphic base T or C

<220>
<221> variation
<222> 2359
<223> polymorphic base A or G

<220>
<221> variation
<222> 2404
<223> polymorphic base A or G

<220>
<221> variation
<222> 2423
<223> polymorphic base T or C

<220>
<221> variation
<222> 2454
<223> polymorphic base T or C

<220>
<221> variation
<222> 2497

<223> polymorphic base A or G

<220>
<221> variation
<222> 2499
<223> polymorphic base A or G

<220>
<221> variation
<222> 2533
<223> polymorphic base T or C

<220>
<221> variation
<222> 2665
<223> polymorphic base T or C

<220>
<221> variation
<222> 2768
<223> insertion of T

<220>
<221> variation
<222> 2855
<223> polymorphic base A or G

<220>
<221> variation
<222> 2858
<223> polymorphic base A or G

<220>
<221> variation
<222> 2867
<223> polymorphic base A or G

<220>
<221> variation
<222> 2870
<223> polymorphic base T or A

<220>
<221> variation
<222> 2874
<223> polymorphic base A or G

<220>
<221> variation
<222> 2881
<223> polymorphic base A or G

<220>
<221> variation
<222> 2882
<223> polymorphic base A or G

<220>
<221> variation
<222> 2898
<223> polymorphic base A or G

<220>

<221> variation
<222> 2910
<223> polymorphic base A or G

<220>
<221> variation
<222> 2933
<223> polymorphic base A or G

<220>
<221> variation
<222> 2946
<223> polymorphic base A or G

<220>
<221> variation
<222> 2957
<223> polymorphic base T or C

<220>
<221> variation
<222> 2961
<223> polymorphic base A or G

<220>
<221> variation
<222> 2981
<223> polymorphic base A or G

<220>
<221> variation
<222> 3001
<223> polymorphic base A or G

<220>
<221> variation
<222> 3006
<223> polymorphic base T or C

<220>
<221> variation
<222> 3015
<223> polymorphic base A or G

<220>
<221> variation
<222> 3027
<223> polymorphic base A or G

<400> 2
gcgccgccag gctcgcaagc accgcgtagg ccagctggcc ggatcccgcc gtctgtc 57
atg gcg gcc ccc atc ctg aaa gat gta gtg gcc tat gtt gaa gtg tgg 105
Met Ala Ala Pro Ile Leu Lys Asp Val Val Ala Tyr Val Glu Val Trp
1 5 10 15
tca tcc aat gga aca gaa aat tat tca aag aca ttt aca aca cag ctt 153
Ser Ser Asn Gly Thr Glu Asn Tyr Ser Lys Thr Phe Thr Thr Gln Leu
20 25 30
gtg gat atg ggg gca aag gtt tca aaa act ttt aac aaa caa gta act 201
Val Asp Met Gly Ala Lys Val Ser Lys Thr Phe Asn Lys Gln Val Thr
35 40 45
cac gtt atc ttc aaa gat ggc tac cag agc act tgg gac aaa gct cag 249
His Val Ile Phe Lys Asp Gly Tyr Gln Ser Thr Trp Asp Lys Ala Gln
50 55 60

```
aag aga ggc gta aag ctc gtt tcg gtg ctc tgg gtk gaa aaa tgc agg      297
Lys Arg Gly Val Lys Leu Val Ser Val Leu Trp Val Glu Lys Cys Arg
65                  70                  75                  80
aca gct gga gca cac att gat gaa tca ttg ttc cct gca gct aat atg      345
Thr Ala Gly Ala His Ile Asp Glu Ser Leu Phe Pro Ala Ala Asn Met
                85                  90                  95
aat gaa cac tta tca agc cta att aaa aaa aaa cgt aaa tgt atg cag      393
Asn Glu His Leu Ser Ser Leu Ile Lys Lys Lys Arg Lys Cys Met Gln
            100                 105                 110
ccc aaa gat ttt aat ttt aaa aca cca gaa aat gat aag aga ttt cag      441
Pro Lys Asp Phe Asn Phe Lys Thr Pro Glu Asn Asp Lys Arg Phe Gln
        115                 120                 125
aag aaa ttt gag aaa atg gct aaa gag cta caa agg caa aaa aca aat      489
Lys Lys Phe Glu Lys Met Ala Lys Glu Leu Gln Arg Gln Lys Thr Asn
    130                 135                 140
cta gat gat gat gta cct att ctc tta ttt gaa tct aat ggt tca tta      537
Leu Asp Asp Asp Val Pro Ile Leu Leu Phe Glu Ser Asn Gly Ser Leu
145                 150                 155                 160
ata tat act ccc aca att gaa att aat agt agt cac cac agc gca atg      585
Ile Tyr Thr Pro Thr Ile Glu Ile Asn Ser Ser His His Ser Ala Met
                165                 170                 175
gag aag aga tta caa gag atg aag gag aaa agg gaa aat ctt tcc ccc      633
Glu Lys Arg Leu Gln Glu Met Lys Glu Lys Arg Glu Asn Leu Ser Pro
            180                 185                 190
acc tct tcc caa atg att cag cag tct cat gat aat cca agt aac tct      681
Thr Ser Ser Gln Met Ile Gln Gln Ser His Asp Asn Pro Ser Asn Ser
        195                 200                 205
ctg tgt gaa gca cct ttg aac att tca cgt gat act ttg tgt tca gat      729
Leu Cys Glu Ala Pro Leu Asn Ile Ser Arg Asp Thr Leu Cys Ser Asp
    210                 215                 220
gaa tac ttt gct ggt ggc tta cac tca tct ttt gat gat ctt tgt gga      777
Glu Tyr Phe Ala Gly Gly Leu His Ser Ser Phe Asp Asp Leu Cys Gly
225                 230                 235                 240
aac tca gga tgt gga aat cag gaa agg aag ttg gaa gga tcc att aat      825
Asn Ser Gly Cys Gly Asn Gln Glu Arg Lys Leu Glu Gly Ser Ile Asn
                245                 250                 255
gac att aaa agt gat gtg tgt att tct tca ctt gta ttg aaa gca aat      873
Asp Ile Lys Ser Asp Val Cys Ile Ser Ser Leu Val Leu Lys Ala Asn
            260                 265                 270
aat att cat tca tca cca tct ttc act cac ctc gat aaa tca agt cct      921
Asn Ile His Ser Ser Pro Ser Phe Thr His Leu Asp Lys Ser Ser Pro
        275                 280                 285
cag aaa ttt ctg agt aat ctt tca aag gaa gaa ata aac ttg caa aka      969
Gln Lys Phe Leu Ser Asn Leu Ser Lys Glu Glu Ile Asn Leu Gln Xaa
    290                 295                 300
aat att gca ggt aaa gta gtc acc cct sac caa aag cag gct gca ggt     1017
Asn Ile Ala Gly Lys Val Val Thr Pro Xaa Gln Lys Gln Ala Ala Gly
305                 310                 315                 320
atg tct cag gag acg ttt gaa gag aag tat cgt ttg tct cct acc tta     1065
Met Ser Gln Glu Thr Phe Glu Glu Lys Tyr Arg Leu Ser Pro Thr Leu
                325                 330                 335
tct tca aca aaa ggc cac ctt ttg ata cat tca aga ccc agg agt tcc     1113
Ser Ser Thr Lys Gly His Leu Leu Ile His Ser Arg Pro Arg Ser Ser
            340                 345                 350
tca gta aag aga aaa aga gta tca cat ggc tcc cat tca cct ccg aag     1161
Ser Val Lys Arg Lys Arg Val Ser His Gly Ser His Ser Pro Pro Lys
        355                 360                 365
gaa aaa tgc aag aga aag agg agc acc agg aga tct atc atg ccg agg     1209
Glu Lys Cys Lys Arg Lys Arg Ser Thr Arg Arg Ser Ile Met Pro Arg
    370                 375                 380
ctg cag ctg tgc agg tcg gaa ggc agg ctg cag cac gtg gcg gga cct     1257
Leu Gln Leu Cys Arg Ser Glu Gly Arg Leu Gln His Val Ala Gly Pro
385                 390                 395                 400
```

```
gcc ctg gag gct ctt agc tgt ggg gag tct tca tat gat gac tat ttt     1305
Ala Leu Glu Ala Leu Ser Cys Gly Glu Ser Ser Tyr Asp Asp Tyr Phe
                405                 410                 415
tca cct gat aat ctt aag gaa agg tat tca gag aat ctt cct cct gaa     1353
Ser Pro Asp Asn Leu Lys Glu Arg Tyr Ser Glu Asn Leu Pro Pro Glu
            420                 425                 430
tct cag ctg cca tca agc cct gct cag ttg agc tgc aga agt ctt tct     1401
Ser Gln Leu Pro Ser Ser Pro Ala Gln Leu Ser Cys Arg Ser Leu Ser
        435                 440                 445
aag aag gag aga aca agc ata ttt gaa atg tct gat ttt tcc tgc gtt     1449
Lys Lys Glu Arg Thr Ser Ile Phe Glu Met Ser Asp Phe Ser Cys Val
    450                 455                 460
ggc aaa aaa acc aga aca gtt gac att acc aat ttc aca gca aaa acc     1497
Gly Lys Lys Thr Arg Thr Val Asp Ile Thr Asn Phe Thr Ala Lys Thr
465                 470                 475                 480
atc tcc agt cct cgg aaa act gga aat ggt gaa ggc cgt gca act tcg     1545
Ile Ser Ser Pro Arg Lys Thr Gly Asn Gly Glu Gly Arg Ala Thr Ser
                485                 490                 495
agt tgc gtg act tct gcc cct gaa gaa gcc cta agg tgt tgt aga cag     1593
Ser Cys Val Thr Ser Ala Pro Glu Glu Ala Leu Arg Cys Cys Arg Gln
            500                 505                 510
gct ggg aaa gaa gac gca tgc cca gag gga aat ggc ttt tct tac acc     1641
Ala Gly Lys Glu Asp Ala Cys Pro Glu Gly Asn Gly Phe Ser Tyr Thr
        515                 520                 525
att gag gac cct gct ctt cca aaa gga cat gat gat gat tta act cct     1689
Ile Glu Asp Pro Ala Leu Pro Lys Gly His Asp Asp Asp Leu Thr Pro
    530                 535                 540
ttg gaa gga agc ctt gaa gaa atg aaa gaa gcg gtt ggt ctg aaa agc     1737
Leu Glu Gly Ser Leu Glu Glu Met Lys Glu Ala Val Gly Leu Lys Ser
545                 550                 555                 560
aca cag aac aaa ggt acc act tcc aaa ata tca aac tcc tct gaa ggc     1785
Thr Gln Asn Lys Gly Thr Thr Ser Lys Ile Ser Asn Ser Ser Glu Gly
                565                 570                 575
gaa gcc cag agt gaa cat gag cca tgt ttt ata gtt gac tgt aac atg     1833
Glu Ala Gln Ser Glu His Glu Pro Cys Phe Ile Val Asp Cys Asn Met
            580                 585                 590
gag acg tct aca gaa gag aag gaa aac tta ccc gga gga tac agt gga     1881
Glu Thr Ser Thr Glu Glu Lys Glu Asn Leu Pro Gly Gly Tyr Ser Gly
        595                 600                 605
agt gtt aaa aat aga cca aca agg cat gat gtt tta gat gac tca tgt     1929
Ser Val Lys Asn Arg Pro Thr Arg His Asp Val Leu Asp Asp Ser Cys
    610                 615                 620
gac ggc ttt aag gac ctc atc aaa cct cat gag gaa ttg aag aaa agt     1977
Asp Gly Phe Lys Asp Leu Ile Lys Pro His Glu Glu Leu Lys Lys Ser
625                 630                 635                 640
ggg aga ggc aaa aag cca aca aga aca tta gtc atg aca agc atg cca     2025
Gly Arg Gly Lys Lys Pro Thr Arg Thr Leu Val Met Thr Ser Met Pro
                645                 650                 655
tct gaa aag cag aat gtc gtc atc cag gtt gtg gat aaa ttg aaa ggc     2073
Ser Glu Lys Gln Asn Val Val Ile Gln Val Val Asp Lys Leu Lys Gly
            660                 665                 670
ttt tca att gca cca gac gtc tgt gag amc acg act cac gtg ctt tcc     2121
Phe Ser Ile Ala Pro Asp Val Cys Glu Xaa Thr Thr His Val Leu Ser
        675                 680                 685
ggg aag cca ctt cgc acc ctg aat gtg ctg ctg gga att gcg cgt ggc     2169
Gly Lys Pro Leu Arg Thr Leu Asn Val Leu Leu Gly Ile Ala Arg Gly
    690                 695                 700
tgc tgg gtt ctc tct tat gat tgg gtg cta tgg tct tta gaa ttg ggt     2217
Cys Trp Val Leu Ser Tyr Asp Trp Val Leu Trp Ser Leu Glu Leu Gly
705                 710                 715                 720
cac tgg att tct gag gag ccg ttc gaa ctg tct cac cac ttc cct gca     2265
His Trp Ile Ser Glu Glu Pro Phe Glu Leu Ser His His Phe Pro Ala
                725                 730                 735
```

```
gct ccc ctg tgc cga agy gag tgc cac ttg tct gca ggg ccg tac cgc     2313
Ala Pro Leu Cys Arg Ser Glu Cys His Leu Ser Ala Gly Pro Tyr Arg
            740                 745                 750
gga acc ctc ttt gcc gac cag cca gyg atg ttt gtc tcg cct gcc agc     2361
Gly Thr Leu Phe Ala Asp Gln Pro Xaa Met Phe Val Ser Pro Ala Ser
        755                 760                 765
agc ccc cca gtg gcc aag ctc tgt gaa cta gtc cac ctg tgc gga ggc     2409
Ser Pro Pro Val Ala Lys Leu Cys Glu Leu Val His Leu Cys Gly Gly
    770                 775                 780
cgg gtc agc caa gtc ccc cgc cag gcc agc atc gtc atc ggg ccc tac     2457
Arg Val Ser Gln Val Pro Arg Gln Ala Ser Ile Val Ile Gly Pro Tyr
785                 790                 795                 800
agc gga aag aag aaa gcm aca gtc aag tat ctg tct gag aaa tgg gtc     2505
Ser Gly Lys Lys Lys Ala Thr Val Lys Tyr Leu Ser Glu Lys Trp Val
                805                 810                 815
tta gat tcc atc acc cag cac aag gtc tgt gcc yct gaa aac tac cta     2553
Leu Asp Ser Ile Thr Gln His Lys Val Cys Ala Xaa Glu Asn Tyr Leu
            820                 825                 830
ttg tca caa tga cagtgacctc actggcctgt ggtgactgca cacagctcgc         2605
Leu Ser Gln  *
        835
aaaactgtct ttggatgttc aaatgagaaa caaaactgtg aagagaagga actggcgtat   2665
acaagatgac ttctgatatc atgtttgcca tgtgttgtgg ttcttaagaa ctcataggtg   2725
actttctgat gactgaatgt ctgtttcaga gacgcttcgg gccttttat ttttatttta   2785
ttttttattt tttgagacgg agtcctgccc tgtttcccag gctggagtgc aatggcacaa   2845
tctcggctca ctgcaacctc cacctcccag gttcaagcga ttctgctgcc tcagcctcct   2905
gagtagctgg gattacagat gtgtgccacc atgcctggct aatttttgta gttttagtag   2965
agacagggtt tcgccatgtt ggccaggctg gtctcaaacg cctgagctca ggtgatctgt   3025
caggcctctt ctatagaatt ccagtctttg tgtcttagtc atgatcataa ttgaaaggtc   3085
acagaacctt tgtcattaga gcacagtact gccaaataaa gaatggaaat tcaatgacat   3145
tgttttatta ctgagaacaa ctagagaact ctgcaagttt cttggcttag actcgatctt   3205
tattaataca ttatctatta ggtaggaaag acatttgtca gctattaagg tgacttttat   3265
ctagcggaga ttcctctctt aaagtaatga aaggagatag gtatgggggg tgttatacag   3325
gataattggt gacatctgag tgtcttactt ctgcaagcct gctttatggt gagcaaagca   3385
tcaccagcaa gtgatcacaa tgtccactgg ccgctttttg cctgccgtcc tcgagatgaa   3445
attggcagtt ggggctgatt cacagaaaca ccgatttgtg gctgagcacg gtggctcaca   3505
cctgtcatcc cagccctttg ggaggctgag gtggacagat cacttgaggt caggagttcg   3565
agaccagcct gaccaacgca gcaaaaccca tctctactaa aaatacaaaa atcagctggg   3625
tgtggtggca cacacctgtg gtcccagctc ctcaggagtc tgaggcagaa gaatcgcttg   3685
aacccaagag gcagaggttg cagtgagcca aggttgcagt gaatcaagat tgctccactg   3745
cactccagcc tgggcaacag agtaactctc cttctcaaat aaataaataa ataaataaga   3805
aaca                                                                3809

<210> 3
<211> 835
<212> PRT
<213> Homo sapiens

<220>
<221> VARIANT
<222> 304
<223> Xaa=Arg or Ile

<220>
<221> VARIANT
<222> 314
<223> Xaa=His or Asp

<220>
<221> VARIANT
<222> 682
<223> Xaa=Thr or Asn
```

<220>
<221> VARIANT
<222> 761
<223> Xaa=Val or Ala

<220>
<221> VARIANT
<222> 828
<223> Xaa=Pro or Ser

<220>
<221> VARIANT
<222> 91
<223> Xaa=Met or Ile

<220>
<221> VARIANT
<222> 306
<223> Xaa=Val or Ala

<220>
<221> VARIANT
<222> 413
<223> Xaa=Pro or Ser

<220>
<221> VARIANT
<222> 528
<223> Xaa=Asp or Gly

<220>
<221> VARIANT
<222> 614
<223> Xaa=Val or Ala

<220>
<221> VARIANT
<222> 677
<223> Xaa=Thr or Asn

<220>
<221> VARIANT
<222> 756
<223> Xaa=Val or Ala

<220>
<221> VARIANT
<222> 758
<223> Xaa=Val or Ala

<220>
<221> VARIANT
<222> 809
<223> Xaa=Lys or Glu

<220>
<221> VARIANT
<222> 821
<223> Xaa=Cys or Arg

<220>

<400> 3

```
Met Ala Ala Pro Ile Leu Lys Asp Val Val Ala Tyr Val Glu Val Trp
1               5                  10                  15
Ser Ser Asn Gly Thr Glu Asn Tyr Ser Lys Thr Phe Thr Thr Gln Leu
            20                  25                  30
Val Asp Met Gly Ala Lys Val Ser Lys Thr Phe Asn Lys Gln Val Thr
        35                  40                  45
His Val Ile Phe Lys Asp Gly Tyr Gln Ser Thr Trp Asp Lys Ala Gln
    50                  55                  60
Lys Arg Gly Val Lys Leu Val Ser Val Leu Trp Val Glu Lys Cys Arg
65                  70                  75                  80
Thr Ala Gly Ala His Ile Asp Glu Ser Leu Phe Pro Ala Ala Asn Met
                85                  90                  95
Asn Glu His Leu Ser Ser Leu Ile Lys Lys Lys Arg Lys Cys Met Gln
            100                 105                 110
Pro Lys Asp Phe Asn Phe Lys Thr Pro Glu Asn Asp Lys Arg Phe Gln
        115                 120                 125
Lys Lys Phe Glu Lys Met Ala Lys Glu Leu Gln Arg Gln Lys Thr Asn
    130                 135                 140
Leu Asp Asp Asp Val Pro Ile Leu Leu Phe Glu Ser Asn Gly Ser Leu
145                 150                 155                 160
Ile Tyr Thr Pro Thr Ile Glu Ile Asn Ser Ser His His Ser Ala Met
                165                 170                 175
Glu Lys Arg Leu Gln Glu Met Lys Glu Lys Arg Glu Asn Leu Ser Pro
            180                 185                 190
Thr Ser Ser Gln Met Ile Gln Gln Ser His Asp Asn Pro Ser Asn Ser
        195                 200                 205
Leu Cys Glu Ala Pro Leu Asn Ile Ser Arg Asp Thr Leu Cys Ser Asp
    210                 215                 220
Glu Tyr Phe Ala Gly Gly Leu His Ser Ser Phe Asp Asp Leu Cys Gly
225                 230                 235                 240
Asn Ser Gly Cys Gly Asn Gln Glu Arg Lys Leu Glu Gly Ser Ile Asn
                245                 250                 255
Asp Ile Lys Ser Asp Val Cys Ile Ser Ser Leu Val Leu Lys Ala Asn
            260                 265                 270
Asn Ile His Ser Ser Pro Ser Phe Thr His Leu Asp Lys Ser Ser Pro
        275                 280                 285
Gln Lys Phe Leu Ser Asn Leu Ser Lys Glu Glu Ile Asn Leu Gln Xaa
    290                 295                 300
Asn Ile Ala Gly Lys Val Val Thr Pro Xaa Gln Lys Gln Ala Ala Gly
305                 310                 315                 320
Met Ser Gln Glu Thr Phe Glu Glu Lys Tyr Arg Leu Ser Pro Thr Leu
                325                 330                 335
Ser Ser Thr Lys Gly His Leu Leu Ile His Ser Arg Pro Arg Ser Ser
            340                 345                 350
Ser Val Lys Arg Lys Arg Val Ser His Gly Ser His Ser Pro Pro Lys
        355                 360                 365
Glu Lys Cys Lys Arg Lys Arg Ser Thr Arg Arg Ser Ile Met Pro Arg
    370                 375                 380
Leu Gln Leu Cys Arg Ser Glu Gly Arg Leu Gln His Val Ala Gly Pro
385                 390                 395                 400
Ala Leu Glu Ala Leu Ser Cys Gly Glu Ser Ser Tyr Asp Asp Tyr Phe
                405                 410                 415
Ser Pro Asp Asn Leu Lys Glu Arg Tyr Ser Glu Asn Leu Pro Pro Glu
            420                 425                 430
Ser Gln Leu Pro Ser Ser Pro Ala Gln Leu Ser Cys Arg Ser Leu Ser
        435                 440                 445
Lys Lys Glu Arg Thr Ser Ile Phe Glu Met Ser Asp Phe Ser Cys Val
    450                 455                 460
Gly Lys Lys Thr Arg Thr Val Asp Ile Thr Asn Phe Thr Ala Lys Thr
465                 470                 475                 480
Ile Ser Ser Pro Arg Lys Thr Gly Asn Gly Glu Gly Arg Ala Thr Ser
                485                 490                 495
Ser Cys Val Thr Ser Ala Pro Glu Glu Ala Leu Arg Cys Cys Arg Gln
```

```
            500                 505                 510
Ala Gly Lys Glu Asp Ala Cys Pro Glu Gly Asn Gly Phe Ser Tyr Thr
        515                 520                 525
Ile Glu Asp Pro Ala Leu Pro Lys Gly His Asp Asp Asp Leu Thr Pro
    530                 535                 540
Leu Glu Gly Ser Leu Glu Glu Met Lys Glu Ala Val Gly Leu Lys Ser
545                 550                 555                 560
Thr Gln Asn Lys Gly Thr Thr Ser Lys Ile Ser Asn Ser Ser Glu Gly
                565                 570                 575
Glu Ala Gln Ser Glu His Glu Pro Cys Phe Ile Val Asp Cys Asn Met
            580                 585                 590
Glu Thr Ser Thr Glu Glu Lys Glu Asn Leu Pro Gly Gly Tyr Ser Gly
        595                 600                 605
Ser Val Lys Asn Arg Pro Thr Arg His Asp Val Leu Asp Asp Ser Cys
    610                 615                 620
Asp Gly Phe Lys Asp Leu Ile Lys Pro His Glu Glu Leu Lys Lys Ser
625                 630                 635                 640
Gly Arg Gly Lys Lys Pro Thr Arg Thr Leu Val Met Thr Ser Met Pro
                645                 650                 655
Ser Glu Lys Gln Asn Val Val Ile Gln Val Val Asp Lys Leu Lys Gly
            660                 665                 670
Phe Ser Ile Ala Pro Asp Val Cys Glu Xaa Thr Thr His Val Leu Ser
        675                 680                 685
Gly Lys Pro Leu Arg Thr Leu Asn Val Leu Leu Gly Ile Ala Arg Gly
    690                 695                 700
Cys Trp Val Leu Ser Tyr Asp Trp Val Leu Trp Ser Leu Glu Leu Gly
705                 710                 715                 720
His Trp Ile Ser Glu Glu Pro Phe Glu Leu Ser His His Phe Pro Ala
                725                 730                 735
Ala Pro Leu Cys Arg Ser Glu Cys His Leu Ser Ala Gly Pro Tyr Arg
            740                 745                 750
Gly Thr Leu Phe Ala Asp Gln Pro Xaa Met Phe Val Ser Pro Ala Ser
        755                 760                 765
Ser Pro Pro Val Ala Lys Leu Cys Glu Leu Val His Leu Cys Gly Gly
    770                 775                 780
Arg Val Ser Gln Val Pro Arg Gln Ala Ser Ile Val Ile Gly Pro Tyr
785                 790                 795                 800
Ser Gly Lys Lys Lys Ala Thr Val Lys Tyr Leu Ser Glu Lys Trp Val
                805                 810                 815
Leu Asp Ser Ile Thr Gln His Lys Val Cys Ala Xaa Glu Asn Tyr Leu
            820                 825                 830
Leu Ser Gln
        835

<210> 4
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> sequencing oligonucleotide PrimerPU

<400> 4
tgtaaaacga cggccagt                                                  18

<210> 5
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> sequencing oligonucleotide PrimerRP
```

<400> 5
caggaaacag ctatgacc 18

INTERNATIONAL SEARCH REPORT

Interr al Application No
PCT/IB 00/01098

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 C12Q1/68 C07K16/18 A61K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K C12Q A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL - EM_GSS 'Online! Entry B88009, 10 March 1998 (1998-03-10) ADAMS, M.D. ET AL.: "RPCI11-16G12.TVB RPCI-11 Homo sapiens genomic clone RPCI-11-16G12, genomic survey sequence." XP002152821 the whole document --- DATABASE EMBL - EMEST_HUM7 'Online! Entry/Acc.no. AI829673, 13 July 1999 (1999-07-13) STRAUSBERG, R.: "wf09b04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2350063 3', mRNA sequence." XP002152822 the whole document --- -/--	1,10, 17-22, 56,58-62 2-5,10, 17-22, 56,58-62

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 November 2000	28/11/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Smalt, R

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL - EMEST_HUM15 'Online! Entry HS611123, Acc.no. R56611, 28 May 1995 (1995-05-28) HILLIER, L. ET AL.: "yg94b12.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:41184 5', mRNA sequence." XP002152823 the whole document	2-4,6, 17-22, 56,58-62
A	WO 99 32644 A (BOUGUELERET LYDIE ;CHUMAKOV ILYA (FR); COHEN DANIEL (FR); GENSET () 1 July 1999 (1999-07-01) the whole document	
A	WO 98 45420 A (DIAGNOCURE INC ;BUSSEMAKERS MARION J G (NL)) 15 October 1998 (1998-10-15) the whole document	
A	WO 99 37986 A (UNIV SOUTHERN CALIFORNIA) 29 July 1999 (1999-07-29) the whole document	
P,X	WO 00 09552 A (GENETICS INST) 24 February 2000 (2000-02-24) whole document, particularly passages relating to seq.ID.157.	2-5,16, 56-62

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 56-61 could be at least partially considered as a mere presentation of information (Rule 39.1.(v) PCT), and claims 59-62 at least partially as a computer program (Rule 39.1.(vi) PCT), the search has been carried out as far as possible in our systematic documentation.

INTERNATIONAL SEARCH REPORT

information on patent family members

Intern ial Application No
PCT/IB 00/01098

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9932644 A	01-07-1999	US 5945522 A	31-08-1999
		AU 1574099 A	12-07-1999
		EP 1052292 A	15-11-2000
		EP 0991770 A	12-04-2000
WO 9845420 A	15-10-1998	AU 7019498 A	30-10-1998
		EP 1007650 A	14-06-2000
WO 9937986 A	29-07-1999	AU 2238499 A	09-08-1999
WO 0009552 A	24-02-2000	AU 5557099 A	06-03-2000

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

The terms "PG-3 gene", when used herein, encompasses genomic, mRNA and cDNA sequences encoding the PG-3 protein, including the untranscribed regulatory regions of the genomic DNA.

The term "heterologous protein", when used herein, is intended to designate any protein or polypeptide other than the PG-3 protein. More particularly, the heterologous protein may be a compound which can be used as a marker in further experiments with a PG-3 regulatory region.

The term "isolated" requires that the material be removed from its original environment (e. g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such a polynucleotide could be part of a vector and/or such a polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. As an example, purification from 0.1 % concentration to 10 % concentration is two orders of magnitude. To illustrate, individual cDNA clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10^4-10^6 fold purification of the native message.

The term "purified" is further used herein to describe a polynucleotide or polynucleotide of the invention which has been separated from other compounds including, but not limited to other polynucleotides or polypeptides (such as the enzymes used in the synthesis of the polynucleotide), carbohydrates, lipids, etc.,. The term "purified" may be used to specify the separation of monomeric polypeptides of the invention from oligomeric forms such as homo- or hetero- dimers, trimers, etc. The term "purified" may also be used to specify the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polypeptide or polynucleotide